

**RELAXATIONS AND APPROXIMATIONS OF CHANCE
CONSTRAINED STOCHASTIC PROGRAMS**

A Thesis
Presented to
The Academic Faculty

by

Weijun Xie

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
August, 2017

Copyright © 2017 by Weijun Xie

RELAXATIONS AND APPROXIMATIONS OF CHANCE CONSTRAINED STOCHASTIC PROGRAMS

Approved by:

Professor Shabbir Ahmed, Advisor
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Professor Santanu S. Dey
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Professor James R. Luedtke
Department of Industrial and Systems
Engineering
University of Wisconsin-Madison

Professor Arkadi Nemirovski
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Professor Alexander Shapiro
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Date Approved: 16 May 2017

*To my wife,
Rong,
and to our daughter,
Charlotte.*

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest thanks to my advisor Prof. Shabbir Ahmed. He has done everything he could to help me throughout my years at Georgia Tech. I really appreciate him allowing me to exploring research ideas and pursuing them freely. In addition, he was also extremely helpful in my successful job search.

I also want to thank a number of other faculty who helped me a lot during my PhD study. Thanks Prof. Santanu Dey, Prof. Jim Luedtke (from UW Madison), Prof. Arkadi Nemirovsk, Prof. Yanfeng Ouyang (from UIUC) and Dr. Dawn Strickland for their great help with my job search. Thanks Prof. Mohit Singh, Prof. Yongjia Song (from VCU), Prof. Ruiwei Jiang (from UMich) (and also already mentioned Prof. Luedtke) for our collaborative researches. I also would like to thank Prof. Alex Shapiro for serving as my thesis committee member and his great help with my research. Besides, I enjoyed the courses taught by Prof. Robert Foley, Prof. Richard Peng, Prof. Sebastian Pokutta, Prof. Alejandro Toriello, Prof. Craig Tovey, Prof. Santosh Vempala, and would like to thank them all.

Thanks to my fellow students, their support during my PhD study enriches my life and the conversations and discussions with them inspire my research ideas a lot. I want to thank (in alphabetical order): Fang Cao, Yufeng Cao, Fabien Caspani, Qiushi Chen, Xiaolei Fang, Rui Gao, Joshua Hale, Satya Malladi, Xinyu Min, Kevin Ryan, Qianyi Wang, Fan Ye, Xiaowei Yue, Can Zhang, Helin Zhu, Jikai Zou and so many more. If you are reading this still looking for your name, I am terribly sorry for my carelessness.

Finally, I thank my parents for always being supportive. I want to thank my wife Rong for standing with me over the years and half of the thesis is owing to her.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	ix
LIST OF FIGURES	x
SUMMARY	xi
I INTRODUCTION	1
1.1 Chance Constrained Stochastic Programming	1
1.2 CCSP with finite support	2
1.3 Distributionally Robust CCSP	4
1.4 Organization	6
II NONANTICIPATIVE DUALITY, RELAXATIONS, AND FORMULA-	
TIONS FOR CHANCE-CONSTRAINED STOCHASTIC PROGRAMS	7
2.1 Introduction	7
2.2 Basic lower bounds	9
2.2.1 Continuous relaxation	9
2.2.2 Quantile bound	10
2.2.3 Scenario grouping based lower bound	10
2.3 Lagrangian dual bounds	11
2.3.1 Basic nonanticipative dual	12
2.3.2 Quantile based Lagrangian dual	14
2.4 A Sufficient Condition for $v_1^{LD} = v^*$	16
2.5 Strength of Lagrangian dual bounds	19
2.5.1 Comparing v_1^{LD} and $v^C(M)$	20
2.5.2 Comparing v_1^{LD} and v_2^{LD}	21
2.5.3 Comparisons with v^Q	22
2.5.4 Bound comparison summary	23
2.6 Primal formulations for chance-constrained mixed-integer linear programs	24
2.6.1 Primal formulation corresponding to v_1^{LD}	25

2.6.2	Primal formulation corresponding to v_2^{LD}	28
2.7	Decomposition algorithms	32
2.7.1	A heuristic scheme	33
2.7.2	A scenario decomposition algorithm for chance-constrained 0-1 programs	34
2.7.3	Branch-and-cut algorithms based on primal MILP formulations	35
2.8	Numerical illustration	39
2.8.1	Illustration of bound quality on instances with continuous x	40
2.8.2	Illustration of the branch-and-cut algorithm on instances with continuous x	42
2.8.3	Performance on instances with binary x	43
III	ON QUANTILE CUTS AND THEIR CLOSURE FOR CHANCE CON- STRAINED OPTIMIZATION PROBLEMS	47
3.1	Introduction	47
3.2	Quantile cuts and Mixing inequalities	50
3.3	Quantile closure	52
3.4	Convergence of quantile closures	55
3.5	Approximate quantile closure in the polyhedral setting	60
3.5.1	The approximate scheme	60
3.5.2	Approximation error estimation for covering CCPs	62
3.6	Generalized quantile closure	64
3.7	Separation over the first quantile closure	68
3.7.1	Complexity	68
3.7.2	A heuristic separation algorithm	69
3.8	Numerical illustration	71
IV	ON BICRITERIA APPROXIMATION OF CHANCE-CONSTRAINED COVERING PROBLEMS	74
4.1	Introduction	74
4.1.1	Problem setting	74
4.1.2	Contributions	75
4.1.3	Related Literature	76
4.2	Inapproximability	77

4.2.1	CVaR approximation	77
4.2.2	Scenario approximation	78
4.2.3	Heuristic algorithm in [4]	79
4.2.4	Single-criterion inapproximability	80
4.3	Bicriteria approximation	82
4.3.1	The scaling algorithm	83
4.3.2	Analysis	83
4.3.3	Effect of tighter relaxation	85
4.3.4	Numerical illustration	87
4.4	Extension: approximation algorithm for a CCSP with general probability mass functions	89
V	ON DETERMINISTIC REFORMULATIONS OF DISTRIBUTIONALLY ROBUST JOINT CHANCE CONSTRAINED OPTIMIZATION PROBLEMS	92
5.1	Introduction	92
5.1.1	Problem Setting	92
5.1.2	Contributions	93
5.1.3	Connection to existing works	94
5.2	Preliminaries	96
5.2.1	Notations	96
5.2.2	Some standard results	96
5.2.3	ϕ -conjugate functions	98
5.3	Deterministic formulations	101
5.3.1	Direct reformulation	101
5.3.2	Biconvex approximation	104
5.4	Convexity conditions for Z	110
5.4.1	Single uncertain constraint	110
5.4.2	Single moment constraint	111
5.4.3	Linear moment constraints	117
5.4.4	Nonlinear positively homogeneous moment constraints	119
5.5	Binary DRCCP	123
5.5.1	Mixed integer convex formulation	124

5.5.2	Numerical illustration	130
VI	DISTRIBUTIONALLY ROBUST CHANCE CONSTRAINED OPTI-	
	MAL POWER FLOW WITH RENEWABLES: A CONIC REFORMU-	
	LATION	134
6.1	Introduction	134
6.2	Nomenclature	136
6.2.1	Sets	136
6.2.2	Parameters	136
6.2.3	Decision Variables	137
6.3	Model Formulation	137
6.4	Convex Reformulation of Chance Constraints (6.1e) and (6.1f)	139
6.4.1	Approximation by Two Single-sided Chance Constraints	140
6.4.2	Exact Reformulation	141
6.4.3	Quality of approximation of Z by $Z_A(\epsilon/2), Z_A(\epsilon)$	146
6.5	Numerical Illustration	146
VII	OPTIMIZED BONFERRONI APPROXIMATIONS OF DISTRIBUTION-	
	ALLY ROBUST JOINT CHANCE CONSTRAINTS	150
7.1	Introduction	150
7.1.1	Setting	150
7.1.2	Related Literature	151
7.1.3	Contributions	153
7.2	Optimized Bonferroni Approximation	154
7.2.1	Single chance constraint approximations	154
7.2.2	Comparison of Approximation Schemes	156
7.2.3	Exactness of Optimized Bonferroni Approximation	157
7.3	Ambiguity Set Based on the First Two Moments	160
7.4	Ambiguity Set Based on Marginal Distributions	166
7.5	Binary Decision Variables and Moment-based Ambiguity Sets	171
7.6	Extension: Ambiguity Set with One Linking Constraint	175
REFERENCES	181

LIST OF TABLES

1	Bound comparison for multi-dimensional continuous knapsack instances. . .	41
2	Computational times for computing bounds for multi-dimensional continuous knapsack instances.	41
3	Computational results for MIP formulation (2.4) and branch-and-cut algorithm on multi-dimensional continuous knapsack instances.	43
4	Bound comparison for multi-dimensional binary knapsack instances	44
5	Computational times for computing bounds for multi-dimensional binary knapsack instances	45
6	Performance of scenario decomposition and MIP on the multi-dimensional binary knapsack instances	46
7	Performance of CPLEX with and without quantile cuts	73
8	Performance comparison of exact and the heuristic methods	133
9	Maximum probability of violations and total costs among model (6.18), risk neutral model and BCH model	147
10	Computational time comparison among model (6.18), risk neutral model and BCH model	149
11	Examples of Probability Distribution, cdf, and Concave Points	169

LIST OF FIGURES

1	Bound comparison summary	25
2	Illustration of Example 2	59
3	Illustration of Example 3	60
4	Illustration of the approximation error in Example 3	64
5	Illustration of approximation ratios for a CCSP with discrete support. . . .	88
6	Comparison between model (6.18) and BCH model	149
7	Illustration of Example 12	157
8	Illustration of Example 14	165
9	Illustration of Example 15	171
10	Illustration of Example 16 with $\frac{2\epsilon}{\Delta} = 2$ and $I = 2$	180

SUMMARY

A chance constrained stochastic programming (CCSP) problem involves constraints with random parameters that are required to be satisfied with a prespecified probability threshold. Such constraints are used to model reliability requirements in a variety of application areas such as finance, energy, service and manufacturing. Except under very special conditions, chance constraints impart severe nonconvexities making the optimization problem extremely difficult. Moreover, in many cases, the probability distribution of the random parameters is not fully specified giving rise to additional difficulties. This thesis makes several contributions towards alleviating these two difficulties in CCSP.

In the first part of this thesis we consider CCSP problems with finitely supported probability distributions. Such problems can be reformulated as mixed integer programming (MIP) problems. We propose two new efficiently solvable Lagrangian dual problems for these problems, and show that their corresponding primal formulations lead to MIP formulations that can be stronger than traditional formulations. We next study a well-known family of cuts for these problems known as quantile cuts. We show that the closure of the infinite family of all quantile cuts has a finite description, and a recursive application of quantile closure operations recovers the convex hull of the nonconvex chance constrained set in the limit. Furthermore, we show that in the pure integer setting, the convergence is finite. Our final result in this part concerns with approximation algorithms for CCSP. We first prove that CCSP is constant factor inapproximable in general. On the other hand, for CCSP problems involving covering type constraints, we prove a bicriteria approximation result where, by relaxing the required probability threshold by a constant factor, we can provide a constant factor approximation algorithm.

In the second part of the thesis we consider distributionally robust chance constrained

problems (DRCCPs) where the chance constraint is required to hold for all probability distributions of the random constraint parameters from a given ambiguity set. First, we study DRCCPs involving convex nonlinear uncertain constraints and ambiguity sets specified by convex moment constraints. We develop deterministic reformulations of such DRCCPs and identify conditions under which such reformulations are convex. Our results generalize and extend several existing results on convex reformulations of DRCCPs. Next, we apply the proposed reformulation scheme to an optimal power flow problem involving uncertainty stemming from renewable power generation. In particular, we develop a convex programming approach for a distributionally robust chance constrained optimal power flow model that ensures low probability of violating upper and lower limits of a line/bus capacity under a wide family of distributions of uncertain renewable generation. Finally, we study a conservative approximation - referred to as a Bonferroni approximation - of a *joint* chance constraint, i.e. a chance constraint involving a system of multiple uncertain constraints. The Bonferroni approximation scheme uses the union bound to approximate the joint chance constraint by a system of *single* chance constraints, one for each original uncertain constraint and each of whose probability thresholds needs to be appropriately set. We show that such a Bonferroni approximation is exact when the uncertainties are separable across the individual constraints, i.e., each uncertain constraint involves a different set of uncertain parameters and corresponding distribution families. We show that, while in general the optimization over the Bonferroni approximation is NP-hard, there are various sufficient conditions under which it is convex and tractable.

CHAPTER I

INTRODUCTION

1.1 *Chance Constrained Stochastic Programming*

A chance constrained stochastic programming (CCSP) problem involves constraints with random parameters that are required to be satisfied with a prespecified probability threshold. A general formulation is as follows

$$v^* = \min \{f(x) : x \in S, \mathbb{P}\{\boldsymbol{\xi} \in \Xi : x \in X(\boldsymbol{\xi})\} \geq 1 - \epsilon\}, \quad (1.1)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is an objective function, S is a nonempty set defined by deterministic constraints, $\boldsymbol{\xi}$ is a random vector with probability distribution \mathbb{P} and support Ξ , $X(\boldsymbol{\xi}) := \{x : G(x, \boldsymbol{\xi}) \leq 0\}$ is a set defined by uncertain constraints $G(x, \boldsymbol{\xi}) \leq 0$ with a mapping $G : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R}^m$, and $\epsilon \in (0, 1)$ is a given risk parameter. The CCSP (1.1) seeks an optimal decision vector x which minimizes the objective function $f(x)$ subject to the deterministic constraints S and satisfies the uncertain constraints with probability at least $(1 - \epsilon)$. CCSPs date back to [29], and some early studies can be found in [81]. CCSPs have been used to model reliability requirements in many application areas; for example, finance [77], production [14, 127], management [105], supply chain design [99, 106], telecommunication [35], and power system [19, 97].

In general, CCSPs involve several challenges. First, for a given $x \in S$, computing $\mathbb{P}\{\boldsymbol{\xi} \in \Xi : x \in X(\boldsymbol{\xi})\}$ (i.e., checking the feasibility of (1.1)) involves a multivariate integration, which can be hard [54, 74]. Even when feasibility can be checked easily, e.g. when the underlying distribution is finite, the resulting optimization problem is highly nonconvex leading to an NP-hard optimization problem [68, 82]. Finally, in many cases, the probability distribution of the random parameters is not fully specified giving rise to additional difficulties.

This thesis makes several contributions in alleviating some of the difficulties mentioned above. In the first part of this thesis, we consider CCSP with finite distributions and present

new duality based reformulations, develop a theory of a cut, and propose approximation algorithms. The second part of this thesis is concerned with CCSPs where the underlying probability distribution is not fully specified. In particular, we consider distributionally robust chance constrained problems (DRCCPs) where the chance constraint is required to hold for all probability distributions of the random constraint parameters from a given ambiguity set. We provide results on convex reformulations of these problems when the ambiguity set is defined by moment constraints, present an application of such models in power systems, and analyze an approximation approach. The following two sections provide a summary of our main results.

1.2 CCSP with finite support

We first consider CCSPs with finite support, i.e., we assume that the random vector ξ has a finite support with $\Xi = \{\xi^1, \dots, \xi^N\} \subseteq \mathbb{R}^m$, where each $i \in \mathcal{N} := \{1, \dots, N\}$ is referred to as a scenario and p_i denotes its probability mass. We can then rewrite (1.1) as

$$\mathbf{v}^* = \min \left\{ f(x) : x \in S, \sum_{i \in [N]} p_i \mathbb{I}(G(x, \xi^i) \leq 0) \geq 1 - \epsilon \right\}, \quad (1.2)$$

where indicator function $\mathbb{I}(A)$ is 1 if event A is true; 0, otherwise.

By introducing a binary variable for each individual scenario and adding big-M inequalities to model the indicator function, the CCSP (1.2) has a natural mixed integer programming (MIP) formulation (cf. [15, 67, 76]). However, it has been reported in [82] that the MIP formulation has a weak relaxation bound and is difficult to solve. There has been significant recent progress improving the MIP formulation. For example, [82] and [101] developed efficient procedures to tighten the big-M coefficients. Many researchers have also attempted to investigate the valid inequalities for CCSPs. In particular, they derived a relaxation of (1.2) in the form of the well-studied mixing set [45] and added the corresponding mixing inequalities [1, 56, 66, 68, 133]. Besides exact solution approaches, approximation algorithms of CCSPs (1.2) have also been extensively studied. In [73], the authors developed convex restrictions of the CCSP feasible region, optimizing over which provides a feasible solution. A similar approach in [3] is used to construct convex relaxations of the CCSP.

The papers [8] and [112] proposed Lagrangian relaxation based heuristics to solve CCSPs. However, none of these general methods has a provable performance guarantee. To the best of our knowledge, all existing approximation algorithms with provable guarantees have been proposed for chance constrained *combinatorial* optimization problem. For instance, [41] proposed constant factor approximation algorithms for chance constrained set covering problems. Subsequently, in [42], the authors developed a fully polynomial time approximation scheme for a chance constrained knapsack problem where the item sizes are drawn from independent normal distributions.

We first propose two new Lagrangian dual problems for (1.2) and develop their associated primal formulations which can be used to exactly compute these dual bounds for chance-constrained linear programs, or a lower bound on them for chance-constrained mixed integer programs. We also propose a new heuristic method and two new exact algorithms that make use of these new bounds to solve these problems to optimality. In our numerical study, we find that for all of our instances, the dual bounds can be quickly computed and demonstrate that heuristic solutions are within 4% of optimal. Our exact algorithms are able to solve more than half of the instances to optimality, although there remain some challenging unsolved instances.

Next, we investigate a type of valid inequalities, called “quantile cuts”, for the MIP formulations of (1.2) [119, 120]. The main contributions are summarized below: (i) the closure of all quantile cuts can be described in a finite conjunctive normal form; (ii) successive application of quantile closure operation achieves the convex hull of the chance constrained problem in the limit; and (iii) in the pure integer setting this convergence is finite, and (iv) separation of quantile cuts is in general NP-hard. For chance constrained mixed integer linear program, we also propose an approximate quantile closure by restricting attention to original problem constraints. We generalize the quantile closure by grouping a number of scenarios together. We propose a heuristic separation algorithm to generate quantile cuts from the first closure, and present numerical studies to demonstrate that these cuts can strengthen the root gaps significantly and help with overall performance.

Our final result in this part concerns with approximation algorithms for (1.2). We first

prove that CCSP is constant factor inapproximable in general. On the other hand, for CCSP problems involving covering type constraints, we prove a bicriteria approximation result where, by relaxing the risk level ϵ by a constant factor, we can provide a constant factor approximation algorithm.

1.3 Distributionally Robust CCSP

In practice, the decision makers often have limited distributional information on ξ , making it challenging to commit to a single \mathbb{P} . As a consequence, the optimal solution to CCSP (1.1) can actually perform poorly if the (true) probability distribution of ξ is different from \mathbb{P} . Hence, a natural alternative to (1.1) is a distributionally robust chance constrained program (DRCCP) of the form

$$v^* = \min \left\{ f(x) : x \in S, \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \{ \xi \in \Xi : G(x, \xi) \leq 0 \} \geq 1 - \epsilon \right\}, \quad (1.3)$$

where \mathcal{P} denotes an ambiguity set of probability measures \mathbb{P} on the space Ξ with a sigma algebra \mathcal{F} . In (1.3), we seek a decision vector x to minimize an objective function $f(x)$ subject to a set of deterministic constraints defined by S , and a chance constraint that is required to hold for any probability distribution from the ambiguity set \mathcal{P} with a probability of $1 - \epsilon$. Recall that (1.3) involves a system of uncertain constraints defined by the mapping $G : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R}^m$. When there is a single uncertain constraint, i.e. $m = 1$, we refer to (1.3) by *single* DRCCP, and when $m > 1$ we refer to (1.3) by *joint* DRCCP. We let Z denote the set of decision vectors x induced by distributionally robust chance constraint, i.e.,

$$Z = \left\{ x : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \{ \xi \in \Xi : G(x, \xi) \leq 0 \} \geq 1 - \epsilon \right\}. \quad (1.4)$$

There has been very significant activities in studying convexity of the set Z under different ambiguity sets (see [24, 46, 47, 51, 123]). For example, for single linear DRCCP (i.e., $m = 1$ and $G(\cdot, \cdot)$ is bilinear in (x, ξ)), the authors in [24, 37], showed that if \mathcal{P} consists of all probability distributions with known first and second moments, then the set Z is second-order cone representable. Various similar convexity results hold for single DRCCP when \mathcal{P} also incorporates other distributional information such as the support of ξ [32], or the unimodality of \mathbb{P} [46, 59]. For distributionally robust *joint* chance constraints, conditions

for convexity of Z are scarce. To the best of our knowledge, [47] provides the first convex reformulation of Z for bilinear mapping $G(\cdot, \cdot)$ and the absence of coefficient uncertainty, i.e. $G(x, \xi) = Ax + B\xi$ for some matrices A, B , when \mathcal{P} is characterized by the mean, a positively homogeneous dispersion measure, and a conic support Ξ of ξ . We are not aware of convexity result of joint DRCCP with general mapping $G(\cdot, \cdot)$ (for example, $G(\cdot, \cdot)$ is a bilinear mapping with coefficient uncertainty).

We first study distributionally robust chance constrained problems (DRCCP) with joint nonlinear uncertain constraints under convex moment ambiguity sets. We show that a DRCCP can be reformulated as a convex program if one the following conditions hold: (i) there is a single uncertain constraint, (ii) the ambiguity set is defined by a single moment constraint, (iii) the ambiguity set is defined by linear moment constraints, and (iv) the moment constraints are positively homogeneous with respect to uncertain parameters. We further show that if the decision variables are binary then a DRCCP can be reformulated as a deterministic mixed integer convex program.

Next, we study a distributionally robust chance constrained optimal power flow problem with known first and second moments. We propose an exact second order cone program reformulation of this problem. Our numerical study shows the proposed model can be solved efficiently and the results are quite robust even with larger risk parameters.

Finally, we study a conservative approximation - referred to as a Bonferroni approximation - of a joint DRCCP. The Bonferroni approximation scheme uses the union bound to approximate the joint chance constraint by a system of single chance constraints, one for each original uncertain constraint and each of whose probability thresholds needs to be appropriately set. We show that such a Bonferroni approximation is exact when the uncertainties are separable across the individual constraints, i.e., each uncertain constraint involves a different set of uncertain parameters and corresponding distribution families. We show that, while in general the Bonferroni approximation is NP-hard, there are various sufficient conditions under which it is convex and tractable.

1.4 *Organization*

The remainder of this thesis is organized in six chapters. The first three chapters are related to CCSP with finite support. Chapter 2 presents our results on duality based formulations of CCSP. This chapter is based on the paper [4]. Chapter ?? presents the results on quantile cuts for CCSP. This chapter is based on the papers [119] and [120]. Chapter 4 presents approximation results for CCSP. The next three chapters are related to distributional robustness. Chapter 5 presents our results on convex reformulations of joint DRCCP with moment based ambiguity sets and is based on the paper [118]. Chapter 6 presents an application of DRCCP to optimal power flow. It is based on the paper [117]. Finally, Chapter 7 studies Bonferroni approximations of DRCCP and is based on the paper [121]. For ease of readability, each chapter is self-contained.

CHAPTER II

NONANTICIPATIVE DUALITY, RELAXATIONS, AND FORMULATIONS FOR CHANCE-CONSTRAINED STOCHASTIC PROGRAMS

2.1 *Introduction*

We consider chance-constrained stochastic programs (CCSPs) of the form

$$v^* = \min \{f(x) : x \in S, \mathbb{P}[x \in X(\boldsymbol{\xi})] \geq 1 - \epsilon\}. \quad (2.1)$$

with the following assumptions: (i) the random vector $\boldsymbol{\xi}$ has a finite support, i.e., $\Xi = \{\xi^1, \dots, \xi^N\}$, where each $i \in \mathcal{N} := \{1, \dots, N\}$ is referred to as a scenario; (ii) $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a linear function; i.e., $f(x) = c^\top x$ (if $f(x)$ is nonlinear, we can introduce a new variable y , add a new constraint $f(x) \leq y$ into S and change the objective to minimize y); (iii) $X(\xi^i) \subseteq S$ for each $i \in \mathcal{N}$; otherwise, we can replace $X(\xi^i)$ by $X(\xi^i) \cap S$ which yields an equivalent problem; (iv) the feasible region is nonempty. We can then rewrite (2.1) as

$$v^* = \min \left\{ c^\top x : x \in S, \sum_{i \in \mathcal{N}} p_i \mathbb{I}(x \in X^i) \geq 1 - \epsilon \right\}, \quad (2.2)$$

where S is a mixed-integer set (i.e., $S \subseteq \mathbb{R}^{n-r} \times \mathbb{Z}^r$ with $0 \leq r \leq n$), $X^i := X(\xi^i)$, $\mathbb{I}(\cdot)$ is the indicator function and p_i is the probability mass associated with scenario i . Using a binary variable z_i to model the indicator function for each scenario i , we reformulate (2.1) as

$$v^* = \min \left\{ c^\top x : x \in S, z_i = \mathbb{I}(x \in X^i), i \in \mathcal{N}, z \in Z \right\}, \quad (2.3)$$

where

$$Z := \left\{ z \in \{0, 1\}^N : \sum_{i \in \mathcal{N}} p_i z_i \geq 1 - \epsilon \right\}.$$

We assume throughout that S , and therefore X^i , are compact sets for all $i \in \mathcal{N}$. Our results can be directly generalized to the unbounded case when the sets S and X^i for all $i \in \mathcal{N}$ share the same recession cone.

CCSPs were first studied in [28, 29]. In [67], the authors analyzed the use of sample average approximation (SAA) for obtaining statistical bounds for CCSPs. Related results are presented in [22, 23, 76]. Under this framework, the resulting sampled problem with a finite number of scenarios can be formulated as a large-scale mixed-integer program (MIP), by introducing a binary variable for each individual scenario and adding big-M inequalities into the formulation. However, the natural MIP formulation based on big-M inequalities often has a weak linear programming relaxation [25]. Motivated by this drawback, there has been significant recent works investigating the use of MIP techniques for solving CCSPs having a finite number of scenarios. In particular, the mixing structure of CCSPs has been studied in [56, 66, 68, 133]. For CCSPs with special combinatorial structures, [100, 101] introduced problem formulations without the binary scenario variables. Chance-constrained formulations have also been proposed for two-stage [61] and multi-stage settings [126].

In this chapter we introduce two new Lagrangian relaxation techniques for obtaining lower bounds for the CCSPs (2.3). Inspired by the associated Lagrangian dual problems, we also introduce new MIP formulations of (2.3) that yield stronger relaxations than existing formulations. The Lagrangian relaxations we construct are obtained by variable splitting: creating multiple copies of the variables x , which are constrained to be equal to each other, and then constructing a relaxation in which these “nonanticipativity constraints” are relaxed. In stochastic programming this technique is known as dual decomposition, and was firstly introduced by [87], and used in [26] for obtaining strong relaxations of two-stage stochastic integer programs. See also [34, 94] for more results on dual decomposition in the two-stage stochastic programming. To the best of our knowledge, only [112] and [8] have applied Lagrangian relaxation for CCSPs. In [112], both the nonanticipativity constraints and the knapsack constraint $\sum_{i \in \mathcal{N}} p_i z_i \geq 1 - \epsilon$ are relaxed; and in [8], the original problem constraints defining X^i are relaxed within an augmented Lagrangian relaxation framework. In contrast, we do not relax the knapsack constraint and directly work on the original formulation (2.3), leading to relaxation bounds that are better than existing alternatives. Somewhat surprisingly, even though the knapsack constraint – which links scenarios together – is not relaxed, the majority of the work required to solve our proposed Lagrangian

relaxation problems can be still decomposed by scenarios.

The remainder of this chapter is organized as follows. In Section 2.2, we discuss three valid lower bounds, which can be obtained by continuous relaxation, quantile bounding and scenario grouping. We then provide two new Lagrangian dual formulations based on relaxing the nonanticipativity constraints in Section 2.3. We also provide a sufficient condition that the duality gap will vanish for chance constrained binary program in Section 2.4. In Section 2.5, we compare these bounds with the basic lower bounds introduced in Section 2.2. We derive new primal formulations in Section 2.6 that are related to the dual formulations from Section 2.3. In Section 2.7 we present a heuristic and two new exact algorithms to solve CCSPs. Finally, we devote Section 2.8 to computational illustration of the lower bounds and performances of the proposed algorithms.

2.2 Basic lower bounds

We first present three different lower bounds for the CCSPs (2.3). The first two are known results while the third bound presented in Section 2.2.3 is new.

2.2.1 Continuous relaxation

Assume that for each scenario i , X^i is described by an inequality system $G_i(x) \leq 0$ with a mapping $G_i : S \mapsto \mathbb{R}^{m_i}$. Problem (2.3) can then be modeled as the following MIP:

$$v^* = \min_{x,z} \left\{ c^\top x : x \in S, G_i(x) \leq M_i(1 - z_i), \forall i \in \mathcal{N}, z \in Z \right\}, \quad (2.4)$$

where M_i is a vector of big- M parameters such that M_{ij} gives a valid upper bound of $G_{ij}(x)$ for all feasible x in (2.2), that is, $M_{ij} \geq \max\{G_{ij}(x) : x \in S, \sum_{i \in \mathcal{N}} p_i \mathbb{I}(x \in X^i) \geq 1 - \epsilon\}$ for each $j = 1, 2, \dots, m_i$. It is impractical to compute the tightest possible upper bound, since it involves solving another CCSP. Therefore, computationally tractable techniques for deriving upper bounds have been investigated. For example, one may begin by choosing $M_{ij} \geq \sup\{G_{ij}(x) : x \in S \cap X^i\}$ for all $j = 1, 2, \dots, m_i$. This simple bound could then be strengthened using a coefficient strengthening that considers more of the feasible region of (2.4) as in [82, 101] (see Section 2.8). We assume that whenever a strengthened big- M parameter $M'_{ij} < M_{ij}$ is obtained, we include the valid inequality $G_{ij}(x) \leq M'_{ij}$ in S , so

that $S \subseteq \{x \in \mathbb{R}^n : G_i(x) \leq M'_i, \forall i \in \mathcal{N}\}$. We define $v^C(M)$ to be the optimal objective value of the relaxation of (2.4) in which the integrality constraints on the z variables are relaxed (but the integrality constraints on x , if any, are not relaxed). We denote by $\underline{v}^C(M)$ the relaxation obtained by also relaxing integrality constraints on x . Clearly, we have

$$\underline{v}^C(M) \leq v^C(M) \leq v^*.$$

2.2.2 Quantile bound

Another lower bound for problem (2.3) is the so-called quantile bound [101]. We first calculate for each $i \in \mathcal{N}$,

$$\eta_i = \min_x \left\{ c^\top x : x \in X^i \right\}.$$

We then sort these values to obtain a permutation σ of \mathcal{N} such that $\eta_{\sigma_1} \geq \dots \geq \eta_{\sigma_N}$. The quantile bound is then defined as $v^Q = \eta_{\sigma_q}$, where $q = \min\{k \in \mathcal{N} : \sum_{i=1}^k p_{\sigma_i} > \epsilon\}$. Then clearly

$$v^Q \leq v^*$$

because at least one scenario in the scenario set $\{\sigma_1, \dots, \sigma_q\}$ must be satisfied in a feasible solution.

2.2.3 Scenario grouping based lower bound

We partition the scenarios \mathcal{N} into $K < N$ disjoint subsets $\mathcal{N}_j, j \in \mathcal{K} = \{1, \dots, K\}$ where $\sum_{j=1}^K |\mathcal{N}_j| = N$. For each $j \in \mathcal{K}$, we define $\tilde{z}_j = 1$ if $z_i = 1$ for all the scenarios $i \in \mathcal{N}_j$ and 0 otherwise (i.e., $\tilde{z}_j = \min\{z_i : i \in \mathcal{N}_j\}$). For each $j \in \mathcal{K}$, we define $q_j = \min_{i \in \mathcal{N}_j} p_i$. We then define the following scenario grouping model:

$$v^G = \min \left\{ c^\top x : x \in S, \tilde{z}_j = \mathbb{I} \left(x \in \bigcap_{i \in \mathcal{N}_j} X^i \right), j \in \mathcal{K}, \sum_{j \in \mathcal{K}} q_j \tilde{z}_j \geq \sum_{j \in \mathcal{K}} q_j - \epsilon \right\}. \quad (2.5)$$

The following proposition shows that (2.5) is indeed a valid relaxation for (2.3).

Proposition 1. $v^G \leq v^*$.

Proof. Let (x, z) be any feasible solution to (2.3) and let $\tilde{z}_j = \min\{z_i : i \in \mathcal{N}_j\}$ for each $j \in \mathcal{K}$. We show that (\tilde{z}, x) is feasible to (2.5). By construction it holds that $\tilde{z}_j = \mathbb{I}(x \in$

$\bigcap_{i \in \mathcal{N}_j} X^i$) for each $j \in \mathcal{K}$. We first establish the following inequality for each $j \in \mathcal{K}$:

$$q_j \tilde{z}_j \geq \sum_{i \in \mathcal{N}_j} p_i z_i - \left(\sum_{i \in \mathcal{N}_j} p_i - q_j \right). \quad (2.6)$$

Indeed, if $\tilde{z}_j = 0$ this implies that $z_i = 0$ for some $i \in \mathcal{N}_j$, and thus $\sum_{i \in \mathcal{N}_j} p_i z_i \leq \sum_{i \in \mathcal{N}_j} p_i - q_j$ as required. On the other hand, if $\tilde{z}_j = 1$ this implies that $z_i = 1$ for all $i \in \mathcal{N}_j$, and (2.6) follows. Now, summing (2.6) over all $j \in \mathcal{K}$ yields

$$\sum_{j \in \mathcal{K}} q_j \tilde{z}_j \geq \sum_{j \in \mathcal{K}} \sum_{i \in \mathcal{N}_j} p_i z_i - \sum_{j \in \mathcal{K}} \sum_{i \in \mathcal{N}_j} p_i + \sum_{j \in \mathcal{K}} q_j \geq 1 - \epsilon - 1 + \sum_{j \in \mathcal{K}} q_j,$$

which establishes the result. \square

If we scale the knapsack inequality in (2.5) by $(\sum_{j \in \mathcal{K}} q_j)^{-1}$, this problem is again a CCSP, but with $K < N$ scenarios. Thus, any technique for obtaining a lower bound of a CCSP can also be applied to this relaxation. In particular, the quantile bound may be applied, and the resulting scenario grouping based quantile bound may be better than the original quantile bound (see Section 2.8 for an illustration). The dual bounds that we derive in the following sections may also be applied to a grouping-based relaxation.

2.3 Lagrangian dual bounds

We next introduce two Lagrangian dual problems associated with the CCSPs (2.3) obtained by relaxing nonanticipativity constraints. We use the following standard result (cf. [70]) on a primal characterization of the Lagrangian dual.

Theorem 1. *Consider a mathematical program $\min\{f(x) : H(x) \leq h, x \in X\}$, where f, H are convex functions and X is compact, and let*

$$\mathcal{L}^* := \sup_{\lambda \geq 0} \{f(x) + \lambda H(x) - \lambda h : x \in X\}$$

be the Lagrangian dual value. Then

$$\mathcal{L}^* = \inf\{f(x) : x \in \text{conv}(X), H(x) \leq h\},$$

where $\text{conv}(S)$ denotes the convex hull of the set S .

2.3.1 Basic nonanticipative dual

By making copies of the decision variables x , the problem (2.3) can be reformulated as

$$v^* = \min_{x,z} \sum_{i \in \mathcal{N}} p_i c^\top x^i, \quad (2.7a)$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{N}} p_i H_i x^i = h, \quad (2.7b)$$

$$z_i = \mathbb{I}(x^i \in X^i), \quad i \in \mathcal{N}, \quad (2.7c)$$

$$z \in Z, \quad (2.7d)$$

$$x^i \in S, \quad i \in \mathcal{N}, \quad (2.7e)$$

where (2.7b) enforce the nonanticipativity constraints $x^1 = \dots = x^N$. The Lagrangian dual problem obtained by dualizing these nonanticipativity constraints with dual vector λ can be written as:

$$v_1^{LD} = \max_{\lambda} \{ \mathcal{L}_1(\lambda) - \lambda^\top h \}, \quad (2.8)$$

where

$$\mathcal{L}_1(\lambda) := \min_{x,z} \left\{ \sum_{i \in \mathcal{N}} p_i (c^\top x^i + \lambda^\top H_i x^i) : (2.7c) - (2.7e) \right\}. \quad (2.9)$$

Next, observe that for a fixed $z \in Z$, this problem reduces to:

$$\psi(z) := \min_x \left\{ \sum_{i \in \mathcal{N}} p_i (c^\top x^i + \lambda^\top H_i x^i) : (2.7c), (2.7e) \right\}.$$

This problem decomposes by scenario. Let

$$\theta_i(\lambda) = \min_x \left\{ c^\top x + \lambda^\top H_i x : x \in S \right\} \quad (2.10)$$

and

$$\zeta_i(\lambda) = \min_x \left\{ c^\top x + \lambda^\top H_i x : x \in X^i \right\}. \quad (2.11)$$

Note that the feasible region of $\zeta_i(\lambda)$ is included in that of $\theta_i(\lambda)$, so we have that $\zeta_i(\lambda) \geq \theta_i(\lambda)$ for all $i \in \mathcal{N}$. By compactness of S , both θ and ζ are finite valued for all λ . Then, we have

$$\psi(z) = \sum_{i \in \mathcal{N}} p_i (\theta_i(\lambda)(1 - z_i) + \zeta_i(\lambda)z_i) = \sum_{i \in \mathcal{N}} p_i \theta_i(\lambda) + \sum_{i \in \mathcal{N}} p_i (\zeta_i(\lambda) - \theta_i(\lambda))z_i$$

and so

$$\mathcal{L}_1(\lambda) = \sum_{i \in \mathcal{N}} p_i \theta_i(\lambda) + \min_z \left\{ \sum_{i \in \mathcal{N}} p_i (\zeta_i(\lambda) - \theta_i(\lambda)) z_i : z \in Z \right\}. \quad (2.12)$$

Thus, for a fixed λ , the Lagrangian relaxation value $\mathcal{L}_1(\lambda) - \lambda^\top h$ can be calculated by first calculating the values $\theta_i(\lambda)$ and $\zeta_i(\lambda)$ by solving (2.10) and (2.11) separately for each $i \in \mathcal{N}$, and then solving a single-row knapsack problem.

We close this subsection by noting that the dual problem (2.8) can be interpreted as a stochastic program with a mean-risk objective function. Let $\Delta_i(\lambda) = \zeta_i(\lambda) - \theta_i(\lambda)$ for all $i \in \mathcal{N}$ and let $F_{\Delta(\lambda)}(\cdot)$ denote the cumulative distribution function of $\{\Delta_i(\lambda)\}_{i \in \mathcal{N}}$ while the point mass function is $\mathbb{P}(\Delta(\lambda) = \Delta_i(\lambda)) = p_i$ for all $i \in \mathcal{N}$ and $F_{\Delta(\lambda)}^{-1}(t) := \inf\{s \in \mathbb{R} : F_{\Delta(\lambda)}(s) \geq t\}$. Then $\mathbf{VaR}_{1-\epsilon}(\lambda) := F_{\Delta(\lambda)}^{-1}(1 - \epsilon)$ is the $(1 - \epsilon)$ -value at risk of $\Delta(\lambda)$, and the $(1 - \epsilon)$ -conditional value at risk of $\Delta(\lambda)$ is defined as [86]:

$$\mathbf{CVaR}_{1-\epsilon}(\lambda) := \mathbf{VaR}_{1-\epsilon}(\lambda) + \frac{1}{\epsilon} \mathbb{E}[\Delta(\lambda) - \mathbf{VaR}_{1-\epsilon}(\lambda)]_+,$$

where $[\cdot]_+ = \max\{\cdot, 0\}$

Proposition 2. *If all of the scenarios are equally likely, i.e., $p_i = \frac{1}{N}$ for all $i \in \mathcal{N}$, and ϵN is an integer, then the dual problem (2.8) is equivalent to*

$$v_1^{LD} = \max_{\lambda} \{ \mathbb{E}[\zeta(\lambda)] - \epsilon \mathbf{CVaR}_{1-\epsilon}(\lambda) - \lambda^\top h \}$$

Proof. When the scenarios are equally likely, the knapsack problem (2.12) can be solved by sorting the values $\Delta_i(\lambda), i \in \mathcal{N}$ to find $\mathbf{VaR}_{1-\epsilon}(\lambda)$. That is, we claim that $z_i^*(\lambda) = \mathbb{I}(\Delta_i(\lambda) \leq \mathbf{VaR}_{1-\epsilon}(\lambda))$ for all $i \in \mathcal{N}$ solves (2.12). Indeed, $\Delta_i(\lambda) = \zeta_i(\lambda) - \theta_i(\lambda) \geq 0$ for all $i \in \mathcal{N}$. Thus, to solve (2.12) with $p_i = \frac{1}{N}$ for all $i \in \mathcal{N}$, we can simply choose the smallest values of $\Delta_i(\lambda)$ until the cardinality constraint $\sum_{i \in \mathcal{N}} z_i \geq (1 - \epsilon)N$ is satisfied. Therefore, the optimal solution sets $z_i^*(\lambda) = \mathbb{I}(\Delta_i(\lambda) \leq \mathbf{VaR}_{1-\epsilon}(\lambda))$ for all $i \in \mathcal{N}$. If $N\epsilon$ is an integer, then simple calculations show that $\mathcal{L}_1(\lambda)$ can be further simplified as

$$\begin{aligned} \mathcal{L}_1(\lambda) &= \frac{1}{N} \sum_{i \in \mathcal{N}} \zeta_i(\lambda) - \epsilon \left(\mathbf{VaR}_{1-\epsilon}(\lambda) + \frac{1}{N\epsilon} \sum_{i \in \mathcal{N}} [\Delta_i(\lambda) - \mathbf{VaR}_{1-\epsilon}(\lambda)]_+ \right) \\ &= \mathbb{E}[\zeta(\lambda)] - \epsilon \mathbf{CVaR}_{1-\epsilon}(\lambda). \end{aligned}$$

□

Proposition 2 can be extended to more general distributions by appropriately adjusting the definitions of $\mathbf{VaR}_{1-\epsilon}$ and $\mathbf{CVaR}_{1-\epsilon}$ as in [85].

2.3.2 Quantile based Lagrangian dual

The quantile bound in Section 2.2.2 can be interpreted as a relaxation obtained by creating a copy x^i of the variables x for each $i \in \mathcal{N}$, as in the reformulation (2.7), but then instead of using the *weighted average* of the objective values of these copies, the *maximum* objective function value among the enforced scenarios is used. This motivates the following alternative reformulation of (2.3):

$$v^* = \min_{x, y, z} y, \quad (2.13a)$$

$$\text{s.t. } c^\top x^i \leq y, \quad i \in \mathcal{N}, \quad (2.13b)$$

$$\sum_{i \in \mathcal{N}} p_i H_i x^i = h, \quad (2.13c)$$

$$z_i = \mathbb{I}(x^i \in X^i), i \in \mathcal{N}, \quad (2.13d)$$

$$z \in Z, \quad (2.13e)$$

$$x^i \in S, \quad i \in \mathcal{N}, \quad (2.13f)$$

where (2.13c) – (2.13f) are just a restatement of (2.7b) - (2.7e). For a fixed $y \in \mathbb{R}$, we further define the problem:

$$g(y) := y + \min_{x, z} \{0 : (2.13b) - (2.13f)\}. \quad (2.14)$$

Clearly, $g(y) = y$ if (2.13b) - (2.13f) is feasible for this fixed y value, otherwise (2.14) is infeasible, and we use the convention $g(y) = +\infty$ in this case. Then (2.13) can be formulated as:

$$v^* = \min_y \{g(y) : y \in \mathbb{R}\}.$$

Next, for a fixed $y \in \mathbb{R}$, let

$$R(y) = \{\{x^i, z_i\}_{i \in \mathcal{N}} : (2.13b), (2.13d) - (2.13f)\}$$

be the set of feasible solutions to (2.13) in which the nonanticipativity constraints (2.13c) are relaxed, and the variable y is fixed. Also, define

$$\mathcal{L}_2(\lambda, y) = \min_{x, z} \left\{ \sum_{i \in \mathcal{N}} p_i \lambda^\top H_i x^i : \{x^i, z_i\}_{i \in \mathcal{N}} \in R(y) \right\},$$

and finally

$$\omega_2(y) = y + \max_{\lambda} \{ \mathcal{L}_2(\lambda, y) - \lambda^\top h \}. \quad (2.15)$$

We use the notation $\omega_2(y) = +\infty$ to indicate that the maximization problem in (2.15) is unbounded. In fact, as the following proposition shows, the maximization problem either has an optimal objective value that equals zero or is unbounded.

Proposition 3. *There exists $\bar{y} \in \mathbb{R}$ such that*

$$\omega_2(y) = \begin{cases} y, & \text{if } y \geq \bar{y}, \\ \infty, & \text{if } y < \bar{y}. \end{cases} \quad (2.16)$$

Proof. By Theorem 1, $\omega_2(y) = y + \min\{0 : \{x^i, z_i\}_{i \in \mathcal{N}} \in T(y)\}$ where $T(y) = \{\{x^i, z_i\}_{i \in \mathcal{N}} : (2.13c), \{x^i, z_i\}_{i \in \mathcal{N}} \in \text{conv}(R(y))\}$. Thus, $\omega_2(y) = y$ if $T(y) \neq \emptyset$ and $\omega_2(y) = \infty$, otherwise. Next, for y large enough, any feasible solution to (2.3) can be used to construct a feasible point in $T(y)$ (just set all x^i equal to x), and so for y large enough $\omega_2(y) = y$. In addition, since the set S is compact, it follows that for y small enough the set $R(y)$ is empty, and hence $T(y)$ is empty. The result then follows because $T(y_1) \subseteq T(y_2)$ whenever $y_1 \leq y_2$. \square

We now define our second Lagrangian dual problem as:

$$v_2^{LD} = \min_y \{ \omega_2(y) : y \in \mathbb{R} \} = \min_y \{ y : \omega_2(y) = y \}. \quad (2.17)$$

Theorem 2. $v_2^{LD} \leq v^*$.

Proof. This follows because $\omega_2(y) \leq g(y)$ for all $y \in \mathbb{R}$. \square

We next discuss the calculation of v_2^{LD} . First, for a given λ and y , $\mathcal{L}_2(\lambda, y)$ can be calculated by solving for each $i \in \mathcal{N}$,

$$\bar{\theta}_i(\lambda, y) := \min_x \left\{ \lambda^\top H_i x : c^\top x \leq y, x \in S \right\}$$

and

$$\bar{\zeta}_i(\lambda, y) := \min_x \left\{ \lambda^\top H_i x : c^\top x \leq y, x \in X^i \right\}.$$

Then,

$$\mathcal{L}_2(\lambda, y) = \sum_{i \in \mathcal{N}} p_i \bar{\theta}_i(\lambda, y) + \min_z \left\{ \sum_{i \in \mathcal{N}} p_i (\bar{\zeta}_i(\lambda, y) - \bar{\theta}_i(\lambda, y)) z_i : z \in Z \right\}.$$

The above characterization leads to a bisection procedure to obtain a lower bound on v_2^{LD} .

It takes as input an upper bound, U , on the optimal objective value v^* , which can be obtained by any feasible solution to (2.3), and a lower bound L (we show in Section 2.5 that $L = v^Q$ is valid). At each iteration, we consider the candidate value $y = (U + L)/2$, and use a subgradient method with a specified finite termination condition (e.g., an iteration limit) to obtain a lower bound $\underline{\omega}_2(y)$ on $\omega_2(y)$. If $\underline{\omega}_2(y) > y$, then we can update $L = y$, otherwise we update $U = y$. The bisection procedure terminates when the difference between the upper and lower bounds is less than a given tolerance. At any step of the algorithm, L is a valid lower bound on v_2^{LD} .

2.4 A Sufficient Condition for $v_1^{LD} = v^*$

In this section, we will study a sufficient condition which closes the duality gap. Let us define a set $\chi = \{(x, z) \mid x^i \in S \ \forall i, \sum_{i=1}^N p_i A_i x^i = h, \sum_{i=1}^N p_i z_i \geq 1 - \epsilon, z_i \leq \mathbb{I}(x^i \in X^i), z_i \in \{0, 1\} \ \forall i\}$. For each $i \in [N]$, let define set $\tilde{X}^i = \{(x^i, z_i) \mid x^i \in S, z_i \leq \mathbb{I}(x^i \in X^i), z_i \in \{0, 1\}\}$ and a bound ρ_i as

$$\rho_i := \sup \left\{ c^\top x^i \mid (x^i, z_i) \in \text{conv}(\tilde{X}^i) \right\} - \inf \left\{ c^\top x^i \mid (x^i, z_i) \in \text{conv}(\tilde{X}^i) \right\}. \quad (2.18)$$

We further make the following assumptions:

Assumption 1. (i) (2.3) is feasible and $S, X^i \subseteq \{0, 1\}^n, \forall i$;

(ii) Suppose $h \in \mathbb{R}^m$, then for some set $I \subseteq \{1, \dots, N\}$ with $|I| = m + 2$, given a $(\tilde{x}^i, \tilde{z}_i) \in \text{conv}(\tilde{X}^i), \forall i \in I$, there exists $(x^i, z_i) \in \tilde{X}^i$ for each $i \in I$ such that $\sum_{i \in I} p_i A_i x^i = \sum_{i \in I} p_i A_i \tilde{x}^i, \forall i \in I, \sum_{i \in I} p_i z_i \geq \sum_{i \in I} p_i \tilde{z}_i$.

Note that under Assumption 1(i), the nonanticipativity constraint can be written as

$$x^1 = \sum_{i=1}^N p_i x^i; \quad (2.19)$$

i.e., if $x_j^1 = 1$, then $x_j^i = 1$; otherwise, $x_j^i = 0$. Under this setting, the nonanticipativity constraints have n rows, thus $m = n$.

Proposition 4. *Under Assumption 1, the duality gap is bounded as*

$$v^* - v_1^{LD} \leq (m+2)p_{\max}\rho_{\max}, \quad (2.20)$$

where $\rho_{\max} = \sup_{i=1,\dots,N} \rho_i$, $p_{\max} = \sup_{i=1,\dots,N} p_i$.

Proof. Define set

$$Y_i = \{[p_i A_i x^i, p_i c^\top x^i, p_i z_i] \mid (x^i, z_i) \in \tilde{X}^i\} \quad (2.21)$$

with the summation defined as $Y = \sum_{i=1}^N Y_i$. By Assumption 1(ii), $Y_i, \text{conv}(Y_i), i = 1, \dots, N$ and $Y, \text{conv}(Y)$ are compact.

Besides of relaxing the nonanticipativity constraint, we also relax $\sum_{i=1}^N p_i z_i \geq 1 - \epsilon$ with dual variable μ ; i.e.,

$$v_3^{LD} := \min_{x,z} \left\{ \sum_{i=1}^N p_i (c^\top x^i + \lambda^\top A_i x^i - \mu p_i z_i) - \lambda^\top h + \mu(1 - \epsilon) : (x^i, z_i) \in \tilde{X}^i \right\}, \quad (2.22)$$

and clearly, we have $v_3^{LD} \leq v_1^{LD}$

Now by definition, we have

$$v^* = \min\{w \mid \exists(u, v, w) \in Y, u = h, w \geq 1 - \epsilon\} \quad (2.23)$$

while by the duality argument in Theorem 1, we have

$$v_3^{LD} = \min\{v \mid \exists(u, v, w) \in \text{conv}(Y), u = h, w \geq 1 - \epsilon\} \leq v_1^{LD} \quad (2.24)$$

Similar to [16], we refer to Shapley-Folkman theorem here.

Theorem 3. (*Shapley-Folkman theorem*) *Let $Y_i, i = 1, \dots, N$ be a collection of subsets of \mathbb{R}^{m+2} ($h \in \mathbb{R}^m$), then for each $y \in \text{conv}(Y)$, there exists a subset $I(y) \subset \{1, \dots, N\}$ with cardinality at most $m+2$ such that*

$$y \in \left[\sum_{i \notin I(y)} Y_i + \sum_{i \in I(y)} \text{conv}(Y_i) \right].$$

Let $(\bar{u}, \bar{v}, \bar{w}) \in \text{conv}(Y)$ such that

$$v_3^{LD} = \bar{v}, \bar{u} = h, \bar{w} \geq 1 - \epsilon.$$

Then by Shapley-Folkman theorem, there exists a subset $\hat{I} \subset \{1, \dots, N\}$ with $|\hat{I}| = m+2$ and

$$(\hat{u}_i, \hat{v}_i, \hat{w}_i) \in \text{conv}(Y_i), i \in \hat{I}, \quad \hat{z}_i \leq \mathbb{I}(\hat{x}^i \in X^i), \forall i \notin \hat{I},$$

such that

$$\begin{aligned} \sum_{i \notin \hat{I}} p_i A_i \hat{x}^i + \sum_{i \in \hat{I}} \hat{u}_i &= h \\ \sum_{i \notin \hat{I}} p_i f(\hat{x}^i, \hat{z}_i) + \sum_{i \in \hat{I}} \hat{v}_i &= \bar{v} = v_3^{LD} \\ \sum_{i \notin \hat{I}} p_i \hat{z}_i + \sum_{i \in \hat{I}} \hat{w}_i &= \bar{w} \geq 1 - \epsilon. \end{aligned}$$

By Caratheodory theorem, for each $i \in \hat{I}$, there exists $(x_i^1, z_i^1), \dots, (x_i^{m+3}, z_i^{m+3}) \in \tilde{X}^i = \{(x^i, z_i) \mid x^i \in S, z_i \leq \mathbb{I}(x^i \in X^i), z_i \in \{0, 1\}\}$ and $\alpha_i^1, \dots, \alpha_i^{m+3} \in [0, 1]$ such that

$$\begin{aligned} \sum_{j=1}^{m+3} \alpha_i^j &= 1, & \hat{u}_i &= p_i A_i \sum_{j=1}^{m+3} \alpha_i^j x_i^j, \\ \hat{v}_i &= p_i \sum_{j=1}^{m+3} \alpha_i^j f(x_i^j, z_i^j), & \hat{w}_i &= p_i \sum_{j=1}^{m+3} \alpha_i^j z_i^j. \end{aligned}$$

Meanwhile, by the definition of $\tilde{f}_i, \bar{f}_i, \rho_i, \forall i$, we have

$$\hat{v}_i = p_i \sum_{j=1}^{m+3} \alpha_i^j f(x_i^j, z_i^j) \geq p_i \inf \left\{ c^\top x^i \mid (x^i, z_i) \in \text{conv}(\tilde{X}^i) \right\}, \forall i.$$

We now have

$$\begin{aligned} \sum_{i \notin \hat{I}} p_i A_i \hat{x}^i + \sum_{i \in \hat{I}} p_i A_i \left(\sum_{j=1}^{m+3} \alpha_i^j x_i^j \right) &= h \\ \sum_{i \notin \hat{I}} p_i f(\hat{x}^i, \hat{z}_i) + \sum_{i \in \hat{I}} p_i \sum_{j=1}^{m+3} \alpha_i^j f(x_i^j, z_i^j) &= v_3^{LD} \geq \sum_{i \notin \hat{I}} p_i f(\hat{x}^i, \hat{z}_i) \\ &+ \sum_{i \in \hat{I}} p_i \inf \left\{ c^\top x^i \mid (x^i, z_i) \in \text{conv}(\tilde{X}^i) \right\} \end{aligned}$$

$$\sum_{i \notin \widehat{I}} p_i \widehat{z}_i + \sum_{i \in \widehat{I}} p_i \left(\sum_{j=1}^{m+3} \alpha_i^j z_i^j \right) = \bar{w} \geq 1 - \epsilon.$$

By Assumption (ii), for a given $\delta > 0$ and there exists a $(\widehat{x}^i, \widehat{z}_i) \in \tilde{X}^i, \forall i \in \widehat{I}$ with $\sum_{i \in \widehat{I}} p_i A_i \widehat{x}_i = \sum_{i \in \widehat{I}} p_i A_i \left(\sum_{j=1}^{m+3} \alpha_i^j x_i^j \right), \sum_{i \in \widehat{I}} p_i \widehat{z}_i \geq \sum_{i \in \widehat{I}} p_i \left(\sum_{j=1}^{m+3} \alpha_i^j z_i^j \right)$ and

$$\sum_{i \in \widehat{I}} p_i f(\widehat{x}^i, \widehat{z}_i) \leq \sum_{i \in \widehat{I}} p_i \sup \left\{ c^\top x^i \mid (x^i, z_i) \in \text{conv}(\tilde{X}^i) \right\}.$$

Thus, we now have

$$\begin{aligned} \sum_{i=1}^N p_i A_i \widehat{x}^i &= h \\ v^* &\leq \sum_{i=1}^N p_i f(\widehat{x}^i, \widehat{z}_i) \leq v_3^{LD} + \sum_{i \in \widehat{I}} p_i \rho_i \\ \sum_{i=1}^N p_i \widehat{z}_i &\geq 1 - \epsilon. \end{aligned}$$

Thus

$$v^* - v_3^{LD} \leq \sum_{i \in \widehat{I}} p_i \rho_i \leq (m+2) p_{\max} \rho_{\max}.$$

□

A direct application of Proposition 4 is that the duality gap vanishes as the number of scenarios grows to infinity.

Corollary 1. *Suppose both Assumption 1 and the following regularity condition*

$$\sup_{i=1, \dots, \infty} \rho_i < \infty, \lim_{N \rightarrow \infty} \sup_{i=1, \dots, N} m p_i = 0,$$

hold, then the duality gap will be vanished when N goes to infinity; i.e.,

$$\lim_{N \rightarrow \infty} (v^* - v_1^{LD}) = 0. \quad (2.25)$$

2.5 Strength of Lagrangian dual bounds

In this section, we compare the Lagrangian dual bounds developed in Section 2.3 and the basic lower bounds in Section 2.2.

2.5.1 Comparing v_1^{LD} and $v^C(M)$

We first show that v_1^{LD} is no smaller than $v^C(M)$. Let

$$C_M = \left\{ (x, z) : x \in S, G_i(x) \leq M_i(1 - z_i), z_i \in [0, 1], \forall i \in \mathcal{N}, \sum_{i \in \mathcal{N}} p_i z_i \geq 1 - \epsilon \right\}$$

be the feasible region of the continuous relaxation of (2.4) in which the variables z are relaxed to be continuous.

Theorem 4. *Assume the constraint qualification is satisfied, i.e., there exists $(\hat{x}, \hat{z}) \in \text{int}(\text{conv}(C_M))$, where $\text{int}(\cdot)$ denotes the interior of a set. Then,*

$$v_1^{LD} \geq v^C(M).$$

Proof. First observe that the continuous relaxation of (2.4) is a convex program with a linear objective function over the convex hull of the set C_M , which is assumed to satisfy Slater's condition [98]. Therefore, by strong duality, the Lagrangian dual of this convex program in which the nonanticipativity constraints $\sum_{i \in \mathcal{N}} p_i H_i x^i = h$ are relaxed has the optimal value equal to $v^C(M)$. But the Lagrangian relaxation problem used in this Lagrangian dual is identical to that in (2.8) except that the z variables are relaxed to be continuous. The conclusion follows. \square \square

Next we establish a set of sufficient conditions under which $v^C(M)$ is equal to v_1^{LD} .

Proposition 5. *Suppose that $S = \mathbb{R}_+^n$ and for each $i \in \mathcal{N}$, $p_i = \frac{1}{N}$ and $G_i(x) = \bar{G}_i(x) + M_i$, where $\bar{G}_i(tx) \leq t\bar{G}_i(x)$ for all $t \geq 1$ and $\bar{G}_i(0) \leq 0$. Then $v^C(M) = v_1^{LD}$.*

Proof. We only need to show that $v^C(M) \geq v_1^{LD}$. Recall that

$$\begin{aligned} v^C(M) &= \min_{x, z} \left\{ c^\top x : x \in S, G_i(x) \leq M_i(1 - z_i), z_i \in [0, 1], \forall i \in \mathcal{N}, \right. \\ &\quad \left. \sum_{i \in \mathcal{N}} z_i \geq \lceil N(1 - \epsilon) \rceil \right\} \\ &= \min_{x, z} \left\{ c^\top x : x \in S, G_i(x) \leq M_i(1 - z_i), \forall i \in \mathcal{N}, z \in \text{conv}(Z) \right\}. \end{aligned}$$

From Theorem 1 and equation (2.8) we know that

$$v_1^{LD} = \min_{x, z} \left\{ \sum_{i=1}^N p_i c^\top x^i : \{(x^i, z_i)\}_{i \in \mathcal{N}} \in \text{conv}(\{(x^i, z_i)\}_{i \in \mathcal{N}} : x^i \in S, \right.$$

$$G_i(x^i) \leq M_i(1 - z_i), \forall i \in \mathcal{N}, z \in Z\}, \sum_{i \in \mathcal{N}} p_i H_i x^i = h\}.$$

Let $(\hat{\mathbf{x}}, \hat{\mathbf{z}})$, where $\hat{\mathbf{z}} := \{\hat{z}_i\}_{i \in \mathcal{N}}$, be an optimal solution of the continuous relaxation of (2.4). Since $\hat{\mathbf{z}} \in \text{conv}(Z)$, there exists a set of points $\{\mathbf{z}_k\} \in Z$ such that $\hat{\mathbf{z}} = \sum_k \lambda_k \mathbf{z}_k$ with $\sum_k \lambda_k = 1$ and $\lambda_k > 0$. Construct $\mathbf{x}_k^i = \frac{\hat{\mathbf{x}}}{\hat{z}_i} z_{ik}$ for all k and for all $i \in \mathcal{N}$. Note that the operations are well defined since, for each i , $\hat{z}_i = 1$ (or 0) implies $z_{ik} = 1$ (or 0) for all k , and we assume that $0 \cdot \infty = 0$. It follows that

$$\begin{aligned} \mathbf{x}_k^i &\in S, \mathbf{z}_k \in Z, \\ G_i(\mathbf{x}_k^i) &= \bar{G}_i(\mathbf{x}_k^i) + M_i \leq \frac{1}{\hat{z}_i} \bar{G}_i(\hat{\mathbf{x}}) z_{ik} + M_i \leq M_i(1 - z_{ik}), \forall i \in \mathcal{N}, \end{aligned}$$

where the first equality is the definition of $G_i(\cdot)$, the second inequality follows because if $z_{ik} = 0$, then $\mathbf{x}_k^i = 0$, and $\bar{G}_i(0) \leq 0$; otherwise, $\mathbf{x}_k^i = \frac{\hat{\mathbf{x}}}{\hat{z}_i}$ and $\bar{G}_i(tx) \leq t\bar{G}_i(x)$ for all $t \geq 1$; while the last inequality follows since $G_i(\hat{\mathbf{x}}) \leq M_i(1 - \hat{z}_i)$ or equivalently, $\frac{1}{\hat{z}_i} \bar{G}_i(\hat{\mathbf{x}}) \leq -M_i$. Now define $(\mathbf{x}^i, \hat{\mathbf{z}}) = \sum_k \lambda_k (\mathbf{x}_k^i, \mathbf{z}_k)$ and we have $\mathbf{x}^i = \hat{\mathbf{x}}$ for all $i \in \mathcal{N}$. Hence,

$$\begin{aligned} \{(\mathbf{x}^i, \hat{z}_i)\}_{i \in \mathcal{N}} &\in \text{conv} \left(\left\{ \{(x^i, z_i)\}_{i \in \mathcal{N}} : x^i \in S, G_i(x^i) \leq M_i(1 - z_i), \forall i \in \mathcal{N}, \right. \right. \\ &\quad \left. \left. z \in Z \right\} \right) \end{aligned}$$

and $\{\mathbf{x}^i\}_{i \in \mathcal{N}}$ also satisfies the nonanticipativity constraints. Thus $(\hat{\mathbf{x}}, \hat{\mathbf{z}})$ is also feasible to (2.8) implying $v^C(M) \geq v_1^{LD}$. □ □

A large class of problems that satisfy the conditions of Proposition 5 are chance-constrained covering linear programs with equiprobable scenarios [82]:

$$\min_{x, z} \left\{ c^\top x : A^i x \geq b^i z_i, \forall i \in \mathcal{N}, \sum_{i \in \mathcal{N}} z_i \geq (1 - \epsilon)N, x \geq 0, z_i \in \{0, 1\} \forall i \in \mathcal{N} \right\}.$$

where $A_i \in \mathbb{R}_+^{m_i \times n}$, $b_i \in \mathbb{R}_+^{m_i}$ for all $i \in \mathcal{N}$. Recasting the above problem in the form of (2.4) we note that $S = \mathbb{R}_+^n$, $\bar{G}_i(x) = -A^i x$, $M_i = b_i$. Indeed we have $\bar{G}_i(tx) = -tA^i x = t\bar{G}_i(x)$, $\forall t \geq 1$, and $\bar{G}_i = 0$.

2.5.2 Comparing v_1^{LD} and v_2^{LD}

The following theorem compares the strengths of the two different Lagrangian dual bounds v_1^{LD} and v_2^{LD} .

Theorem 5. $v_1^{LD} \leq v_2^{LD}$.

Proof. Let $Q_i := \{(x^i, z_i) : (2.7c), (2.7e)\} = \{(x^i, z_i) : (2.13d), (2.13f)\}$ for all $i \in \mathcal{N}$. The claim follows since

$$\begin{aligned}
v_1^{LD} &= \max_{\lambda} \min_{x, y, z} \left\{ y + \sum_{i \in \mathcal{N}} p_i \lambda^\top H_i x^i - \lambda^\top h : (x^i, z_i) \in Q_i, \forall i \in \mathcal{N}, \right. \\
&\quad \left. z \in Z, y \geq \sum_{i \in \mathcal{N}} p_i c^\top x^i \right\} \\
&\leq \max_{\lambda} \min_{x, y, z} \left\{ y + \sum_{i \in \mathcal{N}} p_i \lambda^\top H_i x^i - \lambda^\top h : (x^i, z_i) \in Q_i, \forall i \in \mathcal{N}, \right. \\
&\quad \left. z \in Z, y \geq c^\top x^i, \forall i \in \mathcal{N} \right\} \\
&\leq \min_y \max_{\lambda} \min_{x, z} \left\{ y + \sum_{i \in \mathcal{N}} p_i \lambda^\top H_i x^i - \lambda^\top h : (x^i, z_i) \in Q_i, \forall i \in \mathcal{N}, \right. \\
&\quad \left. z \in Z, y \geq c^\top x^i, \forall i \in \mathcal{N} \right\} \\
&= v_2^{LD},
\end{aligned}$$

where the first equality follows from the definition of v_1^{LD} ; the second inequality follows since $y \geq \sum_{i \in \mathcal{N}} p_i c^\top x^i$ is an aggregation of the constraints $y \geq c^\top x^i$ for each $i \in \mathcal{N}$; the third inequality follows from the max-min inequality; and the final equality is from the definition of v_2^{LD} .

□

□

2.5.3 Comparisons with v^Q

We now show that v_2^{LD} is at least as strong as the quantile bound v^Q .

Theorem 6. $v^Q \leq v_2^{LD}$.

Proof. First define

$$v_2^R(y) = \begin{cases} y & R(y) \neq \emptyset, \\ +\infty & \text{otherwise.} \end{cases}$$

Observe that $v_2^R(y) = y + \mathcal{L}_2(0, y)$, and so $v_2^R(y) \leq \omega_2(y)$ for all $y \in \mathbb{R}$ because it is obtained by using $\lambda = 0$ in (2.15). Thus, it follows that $v^R := \min\{v_2^R(y) : y \in \mathbb{R}\} \leq v_2^{LD}$. We show that this first bound is identical to the quantile bound, i.e., $v^R = v^Q$.

Recall that

$$R(y) = \{ \{x^i, z_i\}_{i \in \mathcal{N}} : (2.13b), (2.13d) - (2.13f) \}.$$

We first show $R(v^Q) \neq \emptyset$, which implies $v_2^R(v^Q) = v^Q$ and thus $v^R \leq v^Q$. Indeed, let $I^Q = \{\sigma_q, \dots, \sigma_N\}$ be the set of scenarios i that $\eta_i \leq \eta_{\sigma_q} = v^Q$ for all $i \in I^Q$. By the definition of v^Q , $\sum_{i \in I^Q} p_i \geq 1 - \epsilon$. Also, for each $i \in I^Q$, there exists $\bar{x}^i \in X^i$ with $c^\top \bar{x}^i = \eta_i \leq v^Q$. Next, let $\hat{x} \in \arg \min \{c^\top x : x \in S\}$ and observe that $c^\top \hat{x} \leq \eta_i \leq v^Q$ for all $i \in I^Q$. Then, a feasible point of $R(v^Q)$ is obtained by setting $x^i = \bar{x}^i$ for $i \in I^Q$, $x^i = \hat{x}$ for $i \in \mathcal{N} \setminus I^Q$ and setting $z_i = 1$ for $i \in I^Q$ and $z_i = 0$ otherwise.

Now let $y < v^Q$ and let $I(y) := \{i \in \mathcal{N} : \eta_i \leq y\}$. For each scenario $i \in \mathcal{N} \setminus I(y)$ there is no $x^i \in X^i$ with $c^\top x^i \leq y$. By definition of v^Q , it holds that $\sum_{i \in I(y)} p_i < 1 - \epsilon$. Thus, $R(y) = \emptyset$ and $v_2^R(y) = +\infty$. Thus $v^R > y$. As $y < v^Q$ was arbitrary, we conclude that $v^R \geq v^Q$. \square \square

Neither of $v_1^{LD}, v^C(M)$ has a general bound relationship with v^Q . The computational results in Section 2.8 provide examples where the quantile bound v^Q is stronger than v_1^{LD} or $v^C(M)$, while the following example shows that v_1^{LD} or $v^C(M)$ can be stronger than v^Q .

Example 1. Consider a three-scenario instance as follows: $X^1 = \{x \in \mathbb{R}_+^2 : 0.5x_1 + 2x_2 \geq 1\}$, $X^2 = \{x \in \mathbb{R}_+^2 : 2x_1 + 0.5x_2 \geq 1\}$, $X^3 = \{x \in \mathbb{R}_+^2 : x_1 + x_2 \geq 1\}$, and $S = \mathbb{R}_+^2$. Each scenario happens with probability $1/3$, and we let $\epsilon = 1/3, M = 1$. The objective is to minimize $x_1 + x_2$. For this instance, the quantile bound $v^Q = 0.5$, and $v_1^{LD} = v^C(M) = 4/7$, therefore, $v_1^{LD}, v^C(M)$ are stronger lower bounds.

2.5.4 Bound comparison summary

We close this section by noting a set of sufficient conditions under which there is no duality gap.

Proposition 6. Suppose $G_i : S \rightarrow \mathbb{R}_-^{m_i} \cup \mathcal{R}_M^i$ for all $i \in \mathcal{N}$, where $\mathcal{R}_M^i = \{s \in \mathbb{R}^{m_i} : \|s\|_\infty = M\}$ for all $i \in \mathcal{N}$ and $M \in \mathbb{R}_+$. Then we have $v^C(M) = v_1^{LD} = v_2^{LD} = v^*$.

Proof. From Theorem 2 and Theorems 4 and 5, it is sufficient to show that $v^C(M) \geq v^*$.

Suppose that $(\widehat{\mathbf{x}}, \widehat{\mathbf{z}})$ is an optimal solution of the continuous relaxation of (2.4), where $\widehat{\mathbf{z}} := \{\widehat{z}_i\}_{i \in \mathcal{N}}$. We show that $(\widehat{\mathbf{x}}, \lceil \widehat{\mathbf{z}} \rceil)$ is another optimal solution. Indeed, if $\widehat{\mathbf{z}}$ is integral, then we are done. Otherwise, suppose that there is an i' such that $\widehat{z}_{i'} \in (0, 1)$, then by the definition of $G_{i'}(\cdot)$, we must have $G_{i'}(\widehat{\mathbf{x}}) \leq 0$. Thus, $(\widehat{\mathbf{x}}, \lceil \widehat{\mathbf{z}} \rceil)$ is feasible to the continuous relaxation of (2.4) with the optimal value $v^C(M)$. Since $\lceil \widehat{\mathbf{z}} \rceil \in Z$, $(\widehat{\mathbf{x}}, \lceil \widehat{\mathbf{z}} \rceil)$ is also feasible to (2.4). Thus, $v^C(M) \geq v^*$. \square \square

A large class of problems that satisfy the conditions of Proposition 6 are chance-constrained set covering problems [5, 15]:

$$\min_{x, z} \left\{ c^\top x : A^i x \geq b^i z_i, \forall i \in \mathcal{N}, x \in \{0, 1\}^n, z \in Z \right\}, \quad (2.26)$$

where $A^i \in \{0, 1\}^{m_i \times n}$, $b^i \in \{0, 1\}^{m_i}$ and $\|b^i\|_\infty = 1$ for all $i \in \mathcal{N}$. Here $G_i(x) = -A^i x + b^i$ and so $G_i(x) : \{0, 1\}^n \rightarrow \mathbb{Z}_-^{m_i} \cup \mathcal{R}_1^i$ with $\mathcal{R}_1^i = \{s \in \{0, 1\}^{m_i} : \|s\|_\infty = 1\}$, for all $i \in \mathcal{N}$ and hence $v^C(M) = v_1^{LD} = v_2^{LD} = v^*$.

The relationships between the basic lower bounds of Section 2.2 and the Lagrangian dual bounds of Section 2.3 are summarized in Figure 1.

2.6 Primal formulations for chance-constrained mixed-integer linear programs

In this section we consider chance-constrained mixed-integer linear programs (MILP), i.e., problem (2.3) where $S = \{x \in \mathbb{R}^{n-r} \times \mathbb{Z}^r : Dx \leq d\}$ and $X^i = \{x \in \mathbb{R}^{n-r} \times \mathbb{Z}^r : A^i x \leq b^i\}$ for each $i \in \mathcal{N}$. Recall our assumption that, for all $i \in \mathcal{N}$, $X^i \subseteq S$ and so we may assume the constraints $Dx \leq d$ are included in the constraints $A^i x \leq b^i$. We derive two new formulations for such problems that are inspired by the two Lagrangian dual problems proposed in the previous section. In particular, under certain conditions, these relaxations are primal formulations of the Lagrangian duals. The constructions here can be extended to the case where S , X^i and f are defined by convex functions using the perspective function approach in [27].

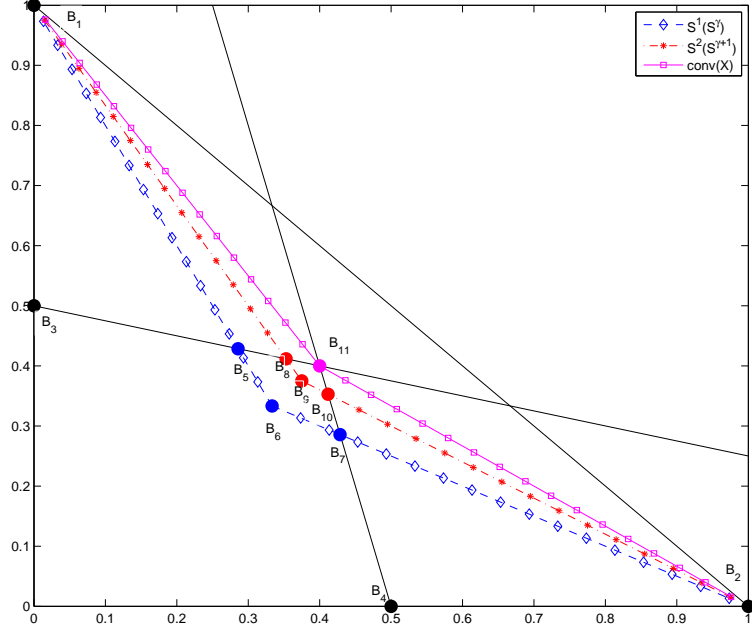


Figure 1: Bound comparison summary

2.6.1 Primal formulation corresponding to v_1^{LD}

Note that replacing $z_i = \mathbb{I}(x^i \in X^i)$ in (2.8) by $z_i \leq \mathbb{I}(x^i \in X^i)$ for each $i \in \mathcal{N}$ yields an equivalent formulation. Recall that $Z = \{z \in \{0, 1\}^N : \sum_{i=1}^N p_i z_i \geq 1 - \epsilon\}$ and let

$$T_1 := \left\{ \{(x^i, z_i)\}_{i \in \mathcal{N}} : Dx^i \leq d, z_i \leq \mathbb{I}(A^i x^i \leq b^i), \right. \\ \left. x^i \in \mathbb{R}^{n-r} \times \mathbb{Z}^r, \forall i \in \mathcal{N}, z \in Z \right\}.$$

From Theorem 1 and equation (2.8) we know that

$$v_1^{LD} = \min_{x, z} \left\{ \sum_{i=1}^N p_i c^\top x^i : \{(x^i, z_i)\}_{i \in \mathcal{N}} \in \text{conv}(T_1), \sum_{i \in \mathcal{N}} p_i H_i x^i = h \right\}. \quad (2.27)$$

Next we use an extended formulation of $\text{conv}(T_1)$ to derive a linear programming relaxation of (2.27) in the following form

$$z_1^{LP} := \min_{x, z, u, w} c^\top x \quad (2.28a)$$

$$\text{s.t. } u^i + w^i = x, \quad \forall i \in \mathcal{N}, \quad (2.28b)$$

$$A^i u^i \leq b^i z_i, \quad \forall i \in \mathcal{N}, \quad (2.28c)$$

$$Dw^i \leq d(1 - z_i), \quad \forall i \in \mathcal{N}, \quad (2.28d)$$

$$z \in \text{conv}(Z). \quad (2.28e)$$

We let $P_S := \{x \in \mathbb{R}^n : Dx \leq d\}$ and $P^i := \{x \in \mathbb{R}^n : A^i x \leq b^i\}$, $i \in \mathcal{N}$ be the continuous relaxations of S, X^i for each $i \in \mathcal{N}$ respectively (the sets are identical in the case $r = 0$). The next theorem shows the relationship between v_1^{LD} and z_1^{LP} .

Theorem 7. $v_1^{LD} \geq z_1^{LP}$. If $P_S = \text{conv}(S)$ and $P^i = \text{conv}(X^i)$ for all $i \in \mathcal{N}$ then $v_1^{LD} = z_1^{LP}$.

Proof. We assume that $P_S = \text{conv}(S)$ and $P^i = \text{conv}(X^i)$ for all $i \in \mathcal{N}$, and show that $v_1^{LD} = z_1^{LP}$. This directly implies that $v_1^{LD} \geq z_1^{LP}$ when $P_S \supseteq \text{conv}(S)$ and $P^i \supseteq \text{conv}(X^i)$.

In the following, for the sake of notational simplicity, we use $(\mathbf{x}, \mathbf{z}, \mathbf{u}, \mathbf{w}) = \{(x^i, z_i, u^i, w^i)\}_{i \in \mathcal{N}}$ and the operations on these vectors will be assumed to be scenario-wise, e.g., $\mathbf{x} \cdot \mathbf{z} := \{x^i z_i\}_{i \in \mathcal{N}}$ and $\mathbf{x}/\mathbf{z} := \{x^i/z_i\}_{i \in \mathcal{N}}$ (here, if $z_i = 0$ the corresponding element is defined to be zero).

(i) Define

$$\bar{T}_1 := \left\{ \{(x^i, z_i)\}_{i \in \mathcal{N}} : Dx^i \leq d, z_i \leq \mathbb{I}(A^i x^i \leq b^i), \forall i \in \mathcal{N}, z \in Z \right\}.$$

We show that $\text{conv}(T_1) = \text{conv}(\bar{T}_1)$. Clearly, $\text{conv}(T_1) \subseteq \text{conv}(\bar{T}_1)$ as $T_1 \subseteq \bar{T}_1$. Hence, we only need to show that $\text{conv}(T_1) \supseteq \text{conv}(\bar{T}_1)$ or equivalently, $\text{conv}(T_1) \supseteq \bar{T}_1$.

Let $(\mathbf{x}, \mathbf{z}) \in \bar{T}_1$. Then for each $i \in \mathcal{N}$, we have $x^i \in P^i$ if $z_i = 1$; $x^i \in P_S$, otherwise. Thus, there exists a finite set of vectors $\{x_{k^i}^i\}_{k^i \in K^i}$ and nonnegative weights $\{\lambda_k^i\}_{k \in K^i}$ such that $\sum_{k \in K^i} \lambda_k^i = 1$, $x^i = \sum_{k^i \in K^i} \lambda_{k^i}^i x_{k^i}^i$ and $Dx_{k^i}^i \leq d, z_i \leq \mathbb{I}(A^i x_{k^i}^i \leq b^i)$, $x_{k^i}^i \in \mathbb{R}^{n-r} \times \mathbb{Z}^r$ for each $k^i \in K^i$. Hence, for each $\{k^i\}_{i \in \mathcal{N}} \in \prod_{i \in \mathcal{N}} K^i$, we have $\{(x_{k^i}^i, z_i)\}_{i \in \mathcal{N}} \in T_1$, and $(\mathbf{x}, \mathbf{z}) \in \text{conv}(\{(x_{k^i}^i, z_i)\}_{i \in \mathcal{N}}, \forall \{k^i\}_{i \in \mathcal{N}} \in \prod_{i \in \mathcal{N}} K^i) \subseteq \text{conv}(T_1)$.

(ii) Next, we define the polyhedron:

$$W_1 := \{(x^i, z_i)\}_{i \in \mathcal{N}} : \exists(u^i, w^i), i \in \mathcal{N}, \text{ s.t. (2.28c) - (2.28e), } u^i + w^i = x^i, \forall i \in \mathcal{N}\}.$$

Then, because the constraints $\sum_{i \in \mathcal{N}} p_i H_i x^i = h$ enforce that all vectors x^i are equal to the same vector, say x , (2.28) can be reformulated as:

$$z_1^{LP} = \min_{x, z} \left\{ \sum_{i=1}^N p_i c^\top x^i : \{(x^i, z_i)\}_{i \in \mathcal{N}} \in W_1, \sum_{i \in \mathcal{N}} p_i H_i x^i = h \right\}.$$

Therefore, it is sufficient to show that $\text{conv}(T_1) = \text{conv}(\bar{T}_1) = W_1$. It is clear that $\bar{T}_1 \subseteq W_1$ and hence $\text{conv}(T_1) = \text{conv}(\bar{T}_1) \subseteq W_1$. Thus, we only need to show $W_1 \subseteq \text{conv}(\bar{T}_1)$.

Let $(\mathbf{x}, \mathbf{z}, \mathbf{u}, \mathbf{w}) \in W_1$. As $\mathbf{z} \in \text{conv}(Z)$, there exists a finite set of vectors $\{\bar{\mathbf{z}}_k\}_{k \in K}$ and nonnegative weights $\{\lambda_k\}_{k \in K}$ such that $\mathbf{z} = \sum_{k \in K} \lambda_k \bar{\mathbf{z}}_k$. Now, for each $k \in K$, define vector $\bar{\mathbf{x}}_k = \bar{\mathbf{z}}_k \cdot (\mathbf{u}/\mathbf{z}) + (\mathbf{1} - \bar{\mathbf{z}}_k) \cdot \mathbf{w}/(\mathbf{1} - \mathbf{z})$. Then, a simple calculation would show that $\sum_{k \in K} \lambda_k \bar{\mathbf{x}}_k = \mathbf{x}$.

The vector $(\bar{\mathbf{x}}_k, \bar{\mathbf{z}}_k)$ satisfies $\bar{\mathbf{z}}_k \in Z$, and for $i \in \mathcal{N}$, if $\bar{z}_{ik} = 0$ then $\bar{x}_k^i = w^i/(1 - z_i) \in \text{conv}(S) = P_S$ from (2.28d) and if $\bar{z}_{ik} = 1$, then $\bar{x}_k^i = u^i/z_i \in \text{conv}(X^i) = P^i$ from (2.28c). Thus, $(\bar{\mathbf{x}}_k, \bar{\mathbf{z}}_k) \in \bar{T}_1$ for each $k \in K$, which directly implies that $(\mathbf{x}, \mathbf{z}) \in \text{conv}(\bar{T}_1)$.

□

It follows from Theorem 7 that $v_1^{LD} = z_1^{LP}$ for chance-constrained linear programs (i.e., when $r = 0$).

When $p_i = 1/N$ for all $i \in \mathcal{N}$ then $\text{conv}(Z) = \{z : \sum_{i \in \mathcal{N}} z_i \geq \lceil (1 - \epsilon)N \rceil, z_i \in [0, 1], i \in \mathcal{N}\}$. For general p_i values a description of $\text{conv}(Z)$ in (2.28e) would require the convex hull of the corresponding knapsack set. Since this is in general intractable, we may replace constraint (2.28e) with a suitable polyhedral relaxation, at the expense of weakening the LP relaxation bound.

Inspired by the above primal formulation, we obtain the following “big- M ” free formulation for chance-constrained mixed-integer linear programs.

Proposition 7.

$$v^* = \min_{x,z,u,w} \{c^\top x : (2.28b) - (2.28d), z \in Z, x \in \mathbb{R}^{n-r} \times \mathbb{Z}^r\}, \quad (2.29)$$

is a valid MILP formulation of (2.3).

2.6.2 Primal formulation corresponding to v_2^{LD}

We next derive a primal formulation for v_2^{LD} under certain conditions. From Theorem 1 and equation (2.17), we have

$$v_2^{LD} = \min_{x,y,z} \left\{ y : \exists \{(x^i, z_i)\}_{i \in \mathcal{N}} \in \text{conv}(R(y)), \sum_{i \in \mathcal{N}} p_i H_i x^i = h \right\}, \quad (2.30)$$

where $R(y) = \{(x^i, z_i)_{i \in \mathcal{N}} : c^\top x^i \leq y, Dx^i \leq d, z_i \in \mathbb{I}(A^i x^i \leq b^i), x^i \in \mathbb{R}^{n-r} \times \mathbb{Z}^r, \forall i \in \mathcal{N}, z \in Z\}$. Next we use an extended formulation of $\text{conv}(R(y))$ to derive the following nonlinear programming formulation of (2.30):

$$z_2^{NLP} := \min_{x,y,z,u,w} y \quad (2.31a)$$

$$\text{s.t. } c^\top u^i \leq y z_i, \quad \forall i \in \mathcal{N}, \quad (2.31b)$$

$$c^\top w^i \leq y(1 - z_i), \quad \forall i \in \mathcal{N}, \quad (2.31c)$$

and (2.28b) – (2.28e).

We define $P_S(y) := \{x \in \mathbb{R}^n : c^\top x \leq y, Dx \leq d\}$ and $P^i(y) := \{x \in \mathbb{R}^n : c^\top x \leq y, A^i x \leq b^i\}, i \in \mathcal{N}$. The next theorem shows the relationship between v_2^{LD} and z_2^{NLP} .

Theorem 8. $v_2^{LD} \geq z_2^{NLP}$. If $P_S(y) = \text{conv}(S \cap \{x : c^\top x \leq y\})$ and $P^i(y) = \text{conv}(X^i \cap \{x : c^\top x \leq y\})$ for all $i \in \mathcal{N}$ and for all $y \in \mathbb{R}$, then $v_2^{LD} = z_2^{NLP}$.

Proof. We assume that $P_S(y) = \text{conv}(S \cap \{x : c^\top x \leq y\})$ and $P^i(y) = \text{conv}(X^i \cap \{x : c^\top x \leq y\})$ for all $i \in \mathcal{N}$ and y , and show that $v_2^{LD} = z_2^{NLP}$. This directly implies that $v_2^{LD} \geq z_2^{NLP}$ as $P_S(y) \supseteq \text{conv}(S \cap \{x : c^\top x \leq y\})$ and $P^i(y) \supseteq \text{conv}(X^i \cap \{x : c^\top x \leq y\})$.

The remainder of the proof is almost identical to that of Theorem 7, so we provide a sketch.

(i) Let us define $\bar{R}(y) := \{\{(x^i, z_i)\}_{i \in \mathcal{N}} : c^\top x^i \leq y, Dx^i \leq d, z_i \leq \mathbb{I}(A^i x^i \leq b^i), \forall i \in \mathcal{N}, z \in Z\}$. Using an approach identical to that in part (i) of the proof of Theorem 7 it can be shown that $\text{conv}(R(y)) = \text{conv}(\bar{R}(y))$ for a given y .

(ii) Next, we define a set $W_2(y) = \{\{(x^i, z_i)\}_{i \in \mathcal{N}} : \exists(u^i, w^i), i \in \mathcal{N}, \text{ s.t. (2.28c)–(2.28e), (2.31b), (2.31c), } w^i = x^i, \forall i \in \mathcal{N}\}$ with a given y . Then, because the constraints $\sum_{i \in \mathcal{N}} p_i H_i x^i = h$ enforce that all vectors x^i are equal to the same vector, say x , (2.31) can be reformulated as:

$$z_2^{NLP} = \min_{x, y, z} \left\{ y : \{(x^i, z_i)\}_{i \in \mathcal{N}} \in W_2(y), \sum_{i \in \mathcal{N}} p_i H_i x^i = h \right\}.$$

Therefore, it is sufficient to show that $\text{conv}(R(y)) = \text{conv}(\bar{R}(y)) = W_2(y)$. The proof of this is similar to part (ii) of the proof of Theorem 7.

□

It follows from Theorem 8 that $v_2^{LD} = z_2^{NLP}$ for chance-constrained linear programs (i.e., when $r = 0$).

Although (2.31) is a nonconvex nonlinear program, it can be solved by bisection on the value of y by observing that the feasible region of (2.31) is nonincreasing over y . Thus, z_2^{NLP} can be calculated by finding the minimum value of y for which the feasible regions of (2.31) is nonempty. For any fixed y , the feasibility problem of (2.31) is a linear program. The disadvantage of solving (2.31) by bisection is that it may be difficult to incorporate such a procedure within a standard linear programming based branch-and-cut algorithm. We therefore propose an iterative scheme that solves a sequence of linear programs that generate progressively better lower bounds for z_2^{NLP} , and eventually converges to z_2^{NLP} .

2.6.2.1 A linear programming based approach for z_2^{NLP}

Let ℓ be a lower bound for z_2^{NLP} (e.g., one can use $\underline{v}^C(M)$). Given such a lower bound ℓ , the nonconvex constraints (2.31b) and (2.31c) can be reformulated with linear constraints, leading to the following formulation:

$$z_2^{LP}(\ell) = \min_{x, y, z, u, w} y, \tag{2.32a}$$

$$\text{s.t. } y \geq c^\top u^i + \ell(1 - z_i), \forall i \in \mathcal{N}, \quad (2.32b)$$

$$y \geq c^\top w^i + \ell z_i, \forall i \in \mathcal{N}, \quad (2.32c)$$

and (2.28b) – (2.28e).

Observe that $z_2^{LP}(\ell)$ is an increasing function of ℓ , and if $\ell \leq z_2^{NLP}$ then $z_2^{LP}(\ell) \leq z_2^{NLP}$. Therefore, if we solve (2.32) iteratively and update ℓ using the optimal objective value, eventually we will converge to some $z_2^{LP}(\bar{\ell}) = \bar{\ell} \leq z_2^{NLP}$. In fact, the value $\bar{\ell}$ will be the same as z_2^{NLP} , since if we replace y and ℓ in (2.32b) and (2.32c) by $\bar{\ell}$, we get the same structure as (2.31b) and (2.31c). We formalize these assertions in the next two results.

Proposition 8. *Let $z_2^{NLP} = \ell^*$, then $z_2^{LP}(\ell^*) = \ell^*$.*

Proof. First, $y = \ell^*$ satisfies all the constraints in (2.32), so $z_2^{LP}(\ell^*) \leq \ell^*$. Suppose $z_2^{LP}(\ell^*) < \ell^*$, then there exists $y^* < \ell^*$ and $\{(u^i, w^i, z_i)\}_{i \in \mathcal{N}}$, which is feasible to (2.28b) - (2.28e) and (2.32b) and (2.32c) being:

$$y^* \geq c^\top u^i + \ell^*(1 - z_i), \forall i \in \mathcal{N},$$

$$y^* \geq c^\top w^i + \ell^* z_i, \forall i \in \mathcal{N}.$$

Thus $y^* \geq c^\top u^i + y^*(1 - z_i)$, $y^* \geq c^\top w^i + y^* z_i$ since $z_i \in [0, 1]$ for each $i \in \mathcal{N}$. So this solution is feasible to (2.31), which is a contradiction. \square

Proposition 9. *Suppose $\ell_0 \leq z_2^{NLP}$, and let $\ell_k = z_2^{LP}(\ell_{k-1})$ for $k \geq 1$. Then $\{\ell_k, k \geq 1\}$ converges to z_2^{NLP} .*

Proof. As $\{\ell_k\}$ is bounded above by z_2^{NLP} , the sequence converges to some $\bar{\ell} \leq z_2^{NLP}$ by the monotone convergence theorem. On the other hand, as $z_2^{LP}(\bar{\ell}) = \bar{\ell}$, there exists $\{(u^i, w^i, z_i)\}_{i \in \mathcal{N}}$ such that this solution with $y = \bar{\ell}$ is feasible to (2.32b) and (2.32c) with $\ell = \bar{\ell}$, and (2.28b) - (2.28e). But this implies that this solution is also feasible to (2.31), and hence $z_2^{NLP} \leq \bar{\ell}$. \square

Similar to the primal formulation (2.29), enforcing integrality constraints on z and any integer constrained x in (2.32) yields an alternative “big- M ” free formulation for general chance-constrained mixed-integer linear programs.

Proposition 10.

$$v^* = \min \{y : x \in \mathbb{R}^{n-r} \times \mathbb{Z}^r, z \in Z, (2.32b), (2.32c), \text{ and } (2.28b) - (2.28d)\}, \quad (2.33)$$

is a valid MILP formulation of (2.3).

Recall that the constraints (2.32b) and (2.32c) depend on a given lower bound ℓ . In our arguments above we required $\ell \leq z_2^{NLP}$ in order to argue that the iterative solution of the linear programming relaxation will converge to z_2^{NLP} . However, any $\ell \leq v^*$ can be used for validity of the formulation (2.33). As examples, one may choose to use the quantile bound v^Q , or z_2^{NLP} obtained by iteratively solving (2.32). In Section 2.7, we develop branch-and-cut decomposition algorithms based on MIP formulations (2.29) and (2.33).

2.6.2.2 A second-order cone programming based approach for z_2^{NLP}

Inspired by the nonlinear program (2.31), we consider the following second order cone programming (SOCP) problem

$$z_2^{SOC}(\ell) := \min_{x,y,z,u,w} y, \quad (2.34a)$$

$$\text{s.t. } (c^\top u^i - \ell z_i)^2 \leq y z_i, \quad \forall i \in \mathcal{N}, \quad (2.34b)$$

$$(c^\top w^i - \ell(1 - z_i))^2 \leq y(1 - z_i), \quad \forall i \in \mathcal{N}, \quad (2.34c)$$

and (2.28b) – (2.28e).

where ℓ is a lower bound on z_2^{NLP} . To see that that (2.34) is indeed an SOCP problem note that $y z_i = \frac{1}{4}(y + z_i)^2 - \frac{1}{4}(y - z_i)^2$ and $y(1 - z_i) = \frac{1}{4}(y + (1 - z_i))^2 - \frac{1}{4}(y - (1 - z_i))^2$. We next relate the values $z_2^{SOC}(\ell)$, z_2^{NLP} , and $z_2^{LP}(\ell)$.

Proposition 11. $z_2^{NLP} \geq \sqrt{z_2^{SOC}(\ell)} + \ell \geq z_2^{LP}(\ell)$ for all $\ell \leq z_2^{NLP}$.

Proof. We first show that $z_2^{NLP} \geq \sqrt{z_2^{SOC}(\ell)} + \ell$. Let $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{w})$ be an optimal solution of (2.31). Consider $\mathbf{y}' = (\mathbf{y} - \ell)^2$. It is clear that $(\mathbf{x}, \mathbf{y}', \mathbf{z}, \mathbf{u}, \mathbf{w})$ satisfies (2.28b) – (2.28e). From (2.31b), (2.31c) and the fact that ℓ is a lower bound and $z_i^2 \leq z_i$ for all $i \in \mathcal{N}$, we have

$$(c^\top u^i - \ell z_i)^2 \leq (\mathbf{y} - \ell)^2 z_i^2 \leq \mathbf{y}' z_i, \quad \forall i \in \mathcal{N},$$

$$(c^\top w^i - \ell(1 - z_i))^2 \leq (\mathbf{y} - \ell)^2(1 - z_i)^2 \leq \mathbf{y}'(1 - z_i), \forall i \in \mathcal{N}.$$

Thus, $(\mathbf{x}, \mathbf{y}', \mathbf{z}, \mathbf{u}, \mathbf{w})$ also satisfies (2.34b) and (2.34c). Hence $z_2^{NLP} \geq \sqrt{z_2^{SOC}(\ell)} + \ell$.

Now we show that $\sqrt{z_2^{SOC}(\ell)} + \ell \geq z_2^{LP}(\ell)$. Let $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{u}, \mathbf{w})$ be an optimal solution of (2.34). Consider $\mathbf{y}' = \sqrt{\mathbf{y}} + \ell$. It is clear that $(\mathbf{x}, \mathbf{y}', \mathbf{z}, \mathbf{u}, \mathbf{w})$ satisfies (2.28b) - (2.28e). From (2.31b), (2.31c) and the fact that ℓ is a lower bound and $z_i \leq 1$ for all $i \in \mathcal{N}$, we have

$$\begin{aligned} (c^\top u^i - \ell z_i)^2 &\leq (\mathbf{y}' - \ell)^2 z_i \leq (\mathbf{y}' - \ell)^2, \forall i \in \mathcal{N}, \\ (c^\top w^i - \ell(1 - z_i))^2 &\leq (\mathbf{y}' - \ell)^2(1 - z_i) \leq (\mathbf{y}' - \ell)^2, \forall i \in \mathcal{N}. \end{aligned}$$

Taking square roots of the above inequalities we see that $(\mathbf{x}, \mathbf{y}', \mathbf{z}, \mathbf{u}, \mathbf{w})$ satisfies (2.32b) and (2.32c). Hence $\sqrt{z_2^{SOC}(\ell)} + \ell \geq z_2^{LP}(\ell)$.

□

Based on the above result we can extend the successive linear programming based approach established in Propositions 8 and 9 to one involving solving successive solutions of the SOCP (2.34). Also similar to (2.33), the SOCP (2.34) after enforcing integrality constraints on z and any integer constrained x variables leads to a “big-M free” mixed-integer SOCP (MISOCP) formulation for the general chance-constrained mixed-integer linear programs.

Proposition 12.

$$(v^* - \ell)^2 = \min \{y : x \in \mathbb{R}^{n-r} \times \mathbb{Z}^r, z \in Z, (2.34b), (2.34c), \text{ and } (2.28b) - (2.28d)\}, \quad (2.35)$$

is a valid MISOCP formulation of (2.3).

2.7 Decomposition algorithms

In this section, we introduce a heuristic algorithm inspired by the bisection procedure for calculating the Lagrangian dual v_2^{LD} and also present two exact algorithms for solving CCSPs (2.3).

2.7.1 A heuristic scheme

The idea of our proposed heuristic algorithm is to use bisection to search for a value ℓ so that fixing $y = \ell$ in (2.13) may yield a feasible solution. Let $X^i = \{x \in \mathbb{R}^{n-r} \times \mathbb{Z}^r : G_i(x) \leq 0\}$ for all $i \in \mathcal{N}$ and let L and U be known initial lower and upper bounds on the optimal value of (2.13). For a fixed $y \in [L, U]$, say $y = (L + U)/2$, we consider the following optimization problem that minimizes the sum of infeasibilities for each scenario:

$$\min_{s, x} \sum_{i \in \mathcal{N}} p_i s_i \quad (2.36a)$$

$$\text{s.t. } G_i(x) \leq s_i \mathbf{e}, \quad i \in \mathcal{N}, \quad (2.36b)$$

$$c^\top x \leq y, \quad x \in S, \quad (2.36c)$$

$$x \in \mathbb{R}^{n-r} \times \mathbb{Z}^r, s \in \mathbb{R}_+^N, \quad (2.36d)$$

where \mathbf{e} is a vector of all 1's. This problem is of the form of a two stage stochastic program with simple recourse and can benefit from specialized decomposition schemes for such problems. Given an optimal solution (\hat{x}, \hat{s}) of (2.36), we check if it is feasible to the original problem (2.13). We set $\hat{z}_i = \mathbb{I}(\hat{s}_i = 0)$ for all $i \in \mathcal{N}$. If $\sum_{i \in \mathcal{N}} p_i \hat{z}_i \geq 1 - \epsilon$, then \hat{x} is feasible to (2.13), and therefore y is a valid upper bound for (2.13). Then we can set $U = y$ and repeat the above steps to find a better feasible solution and hence a better upper bound. On the other hand, if $\sum_{i \in \mathcal{N}} p_i \hat{z}_i < 1 - \epsilon$, we set $L = y$ and repeat the above steps to try to find a feasible solution. The detailed procedure is described in Algorithm 1.

Algorithm 1 A bisection-based heuristic.

- 1: Let $L > -\infty$ and $U < \infty$ be known lower and upper bounds for (2.13), let $\delta > 0$ be the stopping tolerance parameter.
 - 2: **while** $U - L > \delta$ **do**
 - 3: $y \leftarrow (L + U)/2$.
 - 4: Let (\hat{x}, \hat{s}) be an optimal solution of (2.36) and set $\hat{z}_i = \mathbb{I}(\hat{s}_i = 0)$ for all $i \in \mathcal{N}$.
 - 5: **if** $\sum_{i=1}^N p_i \hat{z}_i \geq 1 - \epsilon$ **then**
 - 6: $U \leftarrow y$.
 - 7: **else**
 - 8: $L \leftarrow y$.
 - 9: **end if**
 - 10: **end while**
-

Let v^H denote the solution given by Algorithm 1. We next show that $0 \leq v^H - v^* \leq \delta$

under certain conditions.

Proposition 13. *Suppose $G_i(x) : S \rightarrow \mathbb{R}_-^{m_i} \cup \mathcal{R}_M^i, \forall i \in \mathcal{N}$, where $\mathcal{R}_M^i = \{s \in \mathbb{R}^{m_i} : \|s\|_\infty = M\}$ for all $i \in \mathcal{N}$, $M \in \mathbb{R}_+$. Then Algorithm 1 returns a feasible solution with $v^H - v^* \leq \delta$.*

Proof. First of all, we observe that for any optimal solution $(\mathbf{x}^*, \mathbf{s}^*)$ of (2.36) with a given y , s_i^* is either 0 or M for each $i \in \mathcal{N}$. Indeed, if $G_i(\mathbf{x}^*) \leq 0$, then $s_i^* = 0$; otherwise, the smallest s_i^* that can be chosen is M since $G_i(\mathbf{x}^*) \in \mathcal{R}_M^i$.

Suppose that $(\hat{\mathbf{x}}, \hat{\mathbf{z}})$ is an optimal solution of (2.3). Let $\hat{\mathbf{s}} = M\mathbf{e} - M\hat{\mathbf{z}}$, and $(\hat{\mathbf{x}}, \hat{\mathbf{s}})$ is a feasible solution to (2.36) with any $y \geq v^*$ and $\sum_{i \in \mathcal{N}} p_i \hat{\mathbf{s}} / M \leq \epsilon$. Thus, $v^H \leq v^* + \delta$. □

The conditions of Proposition 13 are identical to those in Proposition 6, and the chance-constrained set covering problems (2.26) satisfy these conditions. Note also that for this problem class with an integer cost vector we can choose $\delta < 1$ and recover an exact optimal solution.

2.7.2 A scenario decomposition algorithm for chance-constrained 0-1 programs

For two-stage stochastic programs in which the first-stage variables are all binary [2] presented a scenario decomposition algorithm that uses the nonanticipative Lagrangian dual of such problems. In this approach, feasible solutions from the scenario subproblems are used to update the upper bound. We describe a simple extension of this method to solve chance-constrained 0-1 programs, which can take advantage of the new Lagrangian dual problems proposed in Section 2.3. Exactly solving the Lagrangian dual problems (2.8) and (2.17) may be challenging in computation. However, the scenario decomposition algorithm remains valid even if the Lagrangian dual multipliers are not optimal. In a practical implementation, we may settle with a lower bound of v_1^{LD} , or v_2^{LD} . For example, we may simply use the quantile bound v^Q , or even a valid lower bound from the scenario grouping based relaxation (2.5).

Algorithm 2 provides a description of the scenario decomposition approach. Finite convergence of Algorithm 2 is an immediate consequence of the following three facts: the lower

Algorithm 2 Scenario decomposition algorithm.

```

1: Let  $UB$  be a known upper bound, let  $LB \leftarrow -\infty$ ,  $E = \emptyset$ , and let  $\delta > 0$  be the stopping
   tolerance parameter.
2: while  $UB - LB > \delta$  do
3:   Calculate a lower bound  $\underline{v}$  for the Lagrangian dual  $v_1^{LD}$  or the quantile-based La-
     grangian dual  $v_2^{LD}$ ;
4:   Collect the optimal solutions  $\hat{x}^i, i \in \mathcal{N}$  that correspond to  $\underline{v}$ , let  $E = \cup_{i \in \mathcal{N}} \hat{x}^i$ .
5:   Update  $LB \leftarrow \max\{LB, \underline{v}\}$ .
6:   for  $x \in E$  do
7:     if  $x$  is feasible and  $c^\top x < UB$  then
8:        $UB \leftarrow c^\top x$ .
9:     end if
10:  end for
11:  Let  $X^i = X^i \setminus E, \forall i \in \mathcal{N}$ .
12: end while

```

bound is nondecreasing; a feasible solution is never excluded if it has not been evaluated; and there are finitely many feasible solutions since for each scenario $i \in \mathcal{N}$, X^i is a finite set. Implementation of the update of the set X^i in line 11 can be accomplished with “no good” cut based on the assumption that all the x variables are binary; see [2] for details.

2.7.3 Branch-and-cut algorithms based on primal MILP formulations

Section 2.6 provided two mixed-integer linear programming formulations (2.29) and (2.33) for (2.3) when the objective function and all constraints are linear. These formulations have a set of variables and constraints for each scenario $i \in \mathcal{N}$, so solving them directly may be time-consuming. We next propose branch-and-cut algorithms to solve these two mixed-integer programs.

We first propose a branch-and-cut algorithm for solving (2.29). Given a solution $(\bar{x}, \bar{z}) \in \mathbb{R}^m \times [0, 1]^N$, checking its feasibility to the LP relaxation of (2.29) is equivalent to checking the existence of $\{(u^i, w^i)\}_{i \in \mathcal{N}}$ that satisfies (2.28b) - (2.28d) with (\bar{x}, \bar{z}) . This can be done by solving the following feasibility problem *separately* for each scenario $i \in \mathcal{N}$, where the dual variables associated with each set of constraints are denoted in parenthesis:

$$f^1(i) = \min_{u, w, \rho} \rho \tag{2.37a}$$

$$\text{s.t. } A^i u^i - \rho e \leq b^i \bar{z}_i, \quad (\gamma) \tag{2.37b}$$

$$Du^i - \rho \mathbf{e} \leq d\bar{z}_i, \quad (\pi_1) \quad (2.37c)$$

$$Dw^i - \rho \mathbf{e} \leq d(1 - \bar{z}_i), \quad (\pi_2) \quad (2.37d)$$

$$u^i + w^i = \bar{x}, \quad (\alpha) \quad (2.37e)$$

$$\rho \geq 0. \quad (2.37f)$$

The form of the feasibility problem (2.37) is chosen to impose a particular normalization on the generated cut, inspired by work on cut-generating linear programs for split cuts in mixed-integer programming [10, 38].

If $f^1(i) = 0$, then no cut is generated for scenario i . Otherwise, let $(\bar{\gamma}, \bar{\pi}_1, \bar{\pi}_2, \bar{\alpha})$ be an optimal dual solution of (2.37). Then $(\bar{x}, \bar{y}, \bar{z})$ is cut off by the following Benders feasibility cut:

$$\bar{\alpha}^\top x + (\bar{\gamma}^\top b_i + \bar{\pi}_1^\top d)z_i + \bar{\pi}_2^\top d(1 - z_i) \leq 0. \quad (2.38)$$

This motivates a branch-and-cut algorithm for solving (2.29) based on Benders decomposition, by treating x and z variables as the first-stage variables, and $\{(u^i, w^i)\}_{i \in \mathcal{N}}$ as the second-stage auxiliary variables. The LP relaxation of the master program is given by:

$$\min_{x, z} c^\top x \quad (2.39a)$$

$$\text{s.t.} \quad \sum_{i=1}^N p_i z_i \geq 1 - \epsilon, \quad (2.39b)$$

$$\pi^k z + \alpha^k x \geq \beta^k, \quad k = 1, \dots, K \quad (2.39c)$$

$$Dx \leq d, \quad (2.39d)$$

$$z \in [0, 1]^N, \quad (2.39e)$$

where (2.39c) are Benders feasibility cuts that have been added in the process of solving the LP relaxation (or throughout the branch-and-cut tree). In each of the Benders cuts (2.39c), obtained from the projection from (2.37), only one scenario variable z_i has a nonzero coefficient. The constraints (2.39d) are actually redundant because the cuts (2.39c) will eventually enforce these. However, including them may help improve convergence of the

cutting plane algorithm for solving the relaxation, and also enables generating additional cuts in the case that x contain some integer restrictions.

Similarly, we develop a branch-and-cut algorithm for solving the mixed-integer program (2.33). Recall that defining (2.33) requires specifying a lower bound ℓ on v^* . We can, for example, use the quantile bound v^Q , or z_2^{NLP} obtained by iteratively solving (2.32). A better lower bound ℓ will yield a better LP relaxation bound. Given a solution $(\bar{x}, \bar{y}, \bar{z}) \in \mathbb{R}^m \times \mathbb{R} \times [0, 1]^N$, checking its feasibility to the LP relaxation of (2.33) is decomposable for each scenario. Treating x, y, z as the first-stage variables, and $\{(u^i, w^i)\}_{i \in \mathcal{N}}$ as the second-stage auxiliary variables, the LP relaxation of the master program can be written as:

$$\min_{x, y, z} y \quad (2.40a)$$

$$\text{s.t. } \sum_{i=1}^N p_i z_i \geq 1 - \epsilon, \quad (2.40b)$$

$$\lambda^k y \geq \beta^k + \rho^k z + \alpha^k x, \quad k = 1, \dots, K \quad (2.40c)$$

$$Dx \leq d \quad (2.40d)$$

$$z \in [0, 1]^N. \quad (2.40e)$$

Let $(\bar{x}, \bar{y}, \bar{z})$ be an optimal solution of (2.40), we check its feasibility to the LP relaxation of (2.33) by solving the following feasibility problem for each scenario $i \in \mathcal{N}$, where the dual variables associated with each set of constraints are denoted in parenthesis.

$$f^2(i) = \min_{u, w, \rho} \rho \quad (2.41a)$$

$$\text{s.t. } c^\top w^i - \rho \leq \bar{y} - \ell \bar{z}_i, \quad (\lambda_1) \quad (2.41b)$$

$$c^\top u^i - \rho \leq \bar{y} - \ell(1 - \bar{z}_i), \quad (\lambda_2) \quad (2.41c)$$

$$A^i u^i - \rho \mathbf{e} \leq b^i \bar{z}_i, \quad (\gamma_1) \quad (2.41d)$$

$$D u^i - \rho \mathbf{e} \leq d \bar{z}_i, \quad (\pi_1) \quad (2.41e)$$

$$D w^i - \rho \mathbf{e} \leq d(1 - \bar{z}_i), \quad (\pi_2) \quad (2.41f)$$

$$u^i + w^i = \bar{x}, \quad (\alpha) \quad (2.41g)$$

$$\rho \geq 0. \quad (2.41h)$$

If $f^2(i) = 0$, then no cut is generated for scenario i . Otherwise, let $(\bar{\lambda}_1, \bar{\lambda}_2, \bar{\gamma}_1, \bar{\pi}_1, \bar{\pi}_2, \bar{\alpha})$ be an optimal dual solution. Then, the following Benders feasibility cut will cut off $(\bar{x}, \bar{y}, \bar{z})$:

$$(\bar{\lambda}_1 + \bar{\lambda}_2)y \leq -\bar{\alpha}^\top x + [\ell \bar{\lambda}_1^\top - \bar{\gamma}_1^\top b^i - \bar{\pi}_1^\top d]z_i + [\ell \bar{\lambda}_2^\top - \bar{\pi}_2^\top d](1 - z_i). \quad (2.42)$$

The lower bound ℓ can be updated in the tree based on the best known global lower bound obtained from the branch-and-cut tree. Once ℓ is updated we can also update the previously generated cuts of the form (2.42) with this new bound, provided the data required to calculate the coefficients in (2.42), $\bar{\lambda}_1, \bar{\lambda}_2, \bar{\gamma}_1^\top b^i + \bar{\pi}_1^\top d$, and $\bar{\pi}_2^\top d$, is saved.

When implementing the branch-and-cut algorithm for solving (2.29) and (2.33), it is not necessary to attempt to generate a cut for every scenario in every round of cut generation. To the contrary, it is possible that a small subset of the scenarios may be most important for generating cuts, and so it makes sense to prioritize generating cuts for scenarios that have yielded cuts in previous rounds.

This decomposition scheme can be incorporated within a branch-and-cut algorithm in which some of the x variables are integer constrained. In this case, branching would be done on both integer constrained x variables and the z variables. Cuts of the form (2.39c) or (2.40c) must be separated if possible any time a solution that satisfies the integrality constraints for both x and z is found, via a lazy cut callback. These cuts can also optionally be added throughout the branch-and-cut tree as user cuts, using some rule for balancing the effort generating the cuts with the bound reduction benefit provided by the cuts.

2.7.3.1 Implementation details of branch-and-cut method

We set the number of threads to one and turn off the CPLEX presolve for the branch-and-cut algorithm. First, we solve (2.32) iteratively using Benders decomposition to get the quantile-based Lagrangian dual bound v_2^{LD} as the lower bound ℓ in (2.41). We use the quantile bound as the initial ℓ . We update this bound by the new optimal objective value of the master problem (2.40) at each iteration. We generate Benders cuts (2.42) iteratively by solving the dual of the feasibility LP (2.41) for each scenario i , and we set the cut violation

threshold to 10^{-3} . At each iteration, we also update the Benders cut coefficients with the updated global lower bound value ℓ . When the quantile-based Lagrangian dual bound v_2^{LD} is obtained, we also obtain a set of Benders cuts. Among these, we include in the master problem (2.40) the cuts that are binding with respect to the relaxation solution in the last iteration.

At the root node, we check feasibility for all scenarios and we add Benders cuts whose cut violations are larger than 10^{-5} . At non-root nodes, we maintain this cut generation effort for integer relaxation solutions to guarantee that integer infeasible solutions are excluded. For fractional relaxation solutions, we apply different efforts for generating Benders cuts according to the depth of the branch-and-bound node, and we limit the number of rounds for cut generation to be one at each node. For nodes whose depths are no more than four, we check feasibility for all scenarios, but we only add up to n most violated Benders cuts, where n is the number of x variables in the model; for nodes whose depths are more than four, we only check feasibility for scenarios that have $\hat{z}_i > 1 - 10^{-3}$, and we only add the most violated Benders cut; for nodes whose depths are more than 10, we only try to generate a Benders cut for a fractional relaxation solution if the depth of the node is divisible by 15.

We update the lower bound ℓ by the best relaxation bound obtained so far, and generate Benders cut using this ℓ by solving (2.41). However, we do not update the coefficients of Benders cuts that have been added to the master problem using this new bound ℓ .

2.8 Numerical illustration

In this section we evaluate the proposed bounds and algorithms on a set of multi-dimensional knapsack instances (1-4-multi with 20 variables and 10 rows in each scenario and 1-6-multi with 39 variables and 5 rows in each scenario) from [101] with two different risk tolerance parameters $\epsilon = 0.1$ and $\epsilon = 0.2$. In both instances (1-4-multi and 1-6-multi), we consider four different scenario sizes: $N \in \{100, 500, 1000, 3000\}$. Under each scenario size (e.g., $N = 100$), we perform five different replications. Since the results among different replications are similar, we report averages over the five replications. We consider two sets of instances: continuous x and binary x . The deterministic feasible set is $S = [0, 1]^n$ for

the instances with continuous x variables and $S = \{0, 1\}^n$ for those with binary x variables. For each scenario $i \in \mathcal{N}$, the feasible set is $X^i := \{x \in S : A^i x \leq b^i\}$, where $A^i \in \mathbb{R}_+^{m \times n}$ and $b^i \in \mathbb{R}_+^m$, and the objective is to minimize $-c^\top x$, where the coefficient vector $c \in \mathbb{R}_+^n$.

2.8.1 Illustration of bound quality on instances with continuous x

We first present a numerical illustration on the strength of the Lagrangian dual bounds on instances with continuous x , i.e., when $S = [0, 1]^n$. We compare the proposed Lagrangian bounds v_1^{LD} and v_2^{LD} with the LP relaxation bound $v^C(M)$ ($\underline{v}^C(M)$ and $v^C(M)$ are identical in this case), quantile bound v^Q , and the optimal objective value v^* . For v_1^{LD} , v_2^{LD} and $v^C(M)$, we report bounds that are obtained with and without strengthening the big- M parameters, using the coefficient strengthening procedure introduced in [101]. Since x is continuous, from Theorems 7 and 8, the Lagrangian dual bounds v_1^{LD} , v_2^{LD} are equal to z_1^{LP} , z_2^{NLP} , respectively. Therefore, we compute v_1^{LD} using (2.28), and when computing v_2^{LD} (z_2^{NLP}), we start with the quantile bound v^Q , then solve the primal LP (2.32) iteratively using Benders decomposition.

In Table 1, we show the optimality gaps of $v^C(M)$, v^Q , v_1^{LD} and v_2^{LD} in the columns labeled accordingly, where optimality gap for a given lower bound LB is defined as $(v^* - LB)/|v^*|$. Under each lower bound, the columns under label ‘With big- M Str.’ provides the bounds obtained if the big- M coefficients have been strengthened, and the columns under label ‘No big- M Str.’ provides the bounds obtained without strengthening big- M coefficients. We also show the optimality gap of the heuristic algorithm described in Section 2.7.1 in column labeled v^H , where, if UB is the objective value of the heuristic solution, this gap is calculated as $(UB - v^*)/|v^*|$. In Table 2, we present the average computation time, in seconds, for obtaining each of these bounds. In addition, the column ‘M-T’ displays the time spent on the big- M strengthening procedure, which is a pre-processing step required for calculating the bounds under the ‘With big- M Str.’ columns. (Thus, e.g., the total time for calculating such a bound is the sum of the ‘M-T’ and the bound time.)

We can see from Table 1 that strengthening big- M parameters significantly improves the bounds given by v_1^{LD} and $v^C(M)$. Without strengthening big- M parameters, bounds given

Table 1: Bound comparison for multi-dimensional continuous knapsack instances.

Instance	ϵ	N	No big- M Str.			With big- M Str.			v^Q	v^H
			$v^C(M)$	v_1^{LD}	v_2^{LD}	$v^C(M)$	v_1^{LD}	v_2^{LD}		
1-4-multi (20,10)	0.1	100	10.1%	7.3%	1.3%	2.3%	2.3%	1.2%	2.4%	0.4%
		500	10.0%	7.0%	1.4%	2.4%	2.4%	1.2%	2.1%	0.2%
		1000	10.0%	7.3%	1.6%	2.5%	2.5%	1.4%	2.5%	0.3%
		3000	9.8%	7.2%	1.7%	2.6%	2.6%	1.5%	2.5%	0.2%
	0.2	100	14.5%	10.5%	1.3%	3.0%	3.0%	1.1%	2.0%	0.6%
		500	14.7%	10.3%	1.4%	3.0%	2.9%	1.3%	2.1%	0.3%
		1000	14.8%	10.7%	1.7%	3.2%	3.2%	1.5%	2.5%	0.3%
		3000	14.4%	10.5%	1.8%	3.2%	3.2%	1.6%	2.6%	0.1%
1-6-multi (39,5)	0.1	100	8.0%	7.4%	2.0%	2.0%	2.0%	1.5%	3.2%	0.6%
		500	8.9%	8.3%	2.5%	2.6%	2.5%	2.0%	3.8%	0.3%
		1000	8.8%	8.3%	2.6%	2.6%	2.6%	2.1%	3.9%	0.4%
		3000	8.7%	8.3%	2.9%	2.8%	2.8%	2.3%	4.3%	0.1%
	0.2	100	11.4%	10.7%	2.1%	2.7%	2.7%	1.8%	3.3%	0.4%
		500	12.4%	11.7%	2.6%	3.4%	3.3%	2.2%	3.5%	0.3%
		1000	12.4%	11.7%	2.9%	3.5%	3.4%	2.4%	4.0%	0.2%
		3000	12.1%	11.6%	3.0%	3.5%	3.5%	2.5%	4.2%	0.1%

Table 2: Computational times for computing bounds for multi-dimensional continuous knapsack instances.

Instance	ϵ	N	No big- M Str.			With big- M Str.			M-T	v^H
			$v^C(M)$	v_1^{LD}	v_2^{LD}	$v^C(M)$	v_1^{LD}	v_2^{LD}		
1-4-multi (20,10)	0.1	100	0.0	0.8	2.6	0.0	0.4	0.8	0.1	0.0
		500	0.2	10.5	26.3	0.0	3.4	6.2	3.7	0.2
		1000	0.5	34.6	83.7	0.2	11.8	21.5	14.6	0.9
		3000	4.1	259.2	683.1	2.0	84.4	151.9	132.2	9.5
	0.2	100	0.0	1.0	3.2	0.0	0.5	1.3	0.1	0.0
		500	0.2	10.9	42.0	0.1	3.6	12.5	3.7	0.3
		1000	0.5	39.8	143.4	0.2	13.7	39.0	14.6	1.2
		3000	4.5	301.8	1338.4	2.3	93.5	292.0	132.2	11.9
1-6-multi (39,5)	0.1	100	0.0	1.0	5.4	0.0	0.5	1.2	0.1	0.0
		500	0.1	8.3	40.5	0.1	3.9	9.1	3.5	0.3
		1000	0.5	31.4	127.1	0.2	12.9	29.7	13.9	0.8
		3000	5.4	222.6	961.2	1.1	84.8	187.5	125.5	7.4
	0.2	100	0.0	1.1	8.5	0.0	0.7	2.2	0.1	0.0
		500	0.1	8.7	62.1	0.1	5.3	18.7	3.5	0.3
		1000	0.5	36.3	224.8	0.3	18.0	55.0	13.9	1.0
		3000	4.7	242.7	1636.0	2.4	122.6	382.6	125.2	9.3

by v_1^{LD} and $v^C(M)$ are rather weak, especially when a higher risk tolerance parameter $\epsilon = 0.2$ is used. With strengthened big- M parameters, the difference between v_1^{LD} and $v^C(M)$ is very small. On the other hand, the bound improvement by strengthening big- M parameters is modest for v_2^{LD} , and v_2^{LD} already gives a tight bound even without strengthening big- M parameters. Overall, the best bounds are obtained by using strengthened big- M parameters and v_2^{LD} . We also find that the heuristic scheme yields a very small optimality gap. For large instances where the exact optimal solution may be challenging to find, one may accept the heuristic solution v^H , when the gap between the lower bound given by v_2^{LD} and the upper bound given by v^H is small enough. From Table 2 we see that we can obtain these strong bounds in a small amount of time. Interestingly, we see that formulations with naive big- M parameters take longer to solve than the ones with strengthened big- M parameters, even after including the time spent on strengthening the big- M parameters. Thus, for these instances, big- M strengthening yields improvements in both computation time and bound.

2.8.2 Illustration of the branch-and-cut algorithm on instances with continuous x

In this section we describe computational experiments using the branch-and-cut approach described in Section 2.7.3 for solving formulation (2.33) on instances with continuous x . For all experiments in our test, we use a time limit of 3600 seconds. We use the heuristic solution obtained by the heuristic algorithm in Section 2.7.1 as a MIP start solution. Further implementation details of the branch-and-cut algorithm are given in the Appendix.

In Table 3 we compare the performance of three computational options: the MIP (2.4) using strengthened big- M parameters (MIP-(2.4)), the branch-and-cut algorithm with strengthened big- M parameters (Benders With big- M), and the branch-and-cut algorithm without strengthened big- M parameters (Benders without big- M). For instances that are not solved to optimality within the time limit, we show in parentheses the number of instances out of five replications that are solved to optimality, and report the average optimality gap. For these instances, we use the number of nodes that have been processed up to the time limit to calculate the average number of nodes, and we use the best lower bound obtained within the time limit to calculate the average root relaxation gap.

Table 3: Computational results for MIP formulation (2.4) and branch-and-cut algorithm on multi-dimensional continuous knapsack instances.

Instances			MIP-(2.4)		Benders w/o big- M		Benders With big- M	
Instance	ϵ	N	AvgT	AvgN	AvgT	AvgN	AvgT	AvgN
1-4-multi (20,10)	0.1	100	0.2	159	2.1	444	0.9	266
		500	52.5	19k	376.6	93k	36.8	35k
		1000	2004.9	356k	0.8%(0)	>210k	0.2%(4)	>788k
		3000	1.8%(0)	>205k	1.6%(0)	>111k	1.3%(0)	>546k
	0.2	100	0.7	664	8.8	1161	1.7	473
		500	452.8	166k	0.5%(1)	>297k	760.4	459k
		1000	0.6%(0)	>610k	1.7%(0)	>145k	0.6%(0)	>1007k
		3000	2.5%(0)	>180k	2.8%(0)	>65k	1.9%(0)	>344k
1-6-multi (39,5)	0.1	100	0.3	350	5.4	2231	1.4	917
		500	781.1	393k	0.2%(1)	>383k	339.3	263k
		1000	0.5%(0)	>905k	1.3%(0)	>195k	0.6%(0)	>1237k
		3000	2.1%(0)	>250k	1.7%(0)	>96k	1.4%(0)	>412k
	0.2	100	1.2	2294	31.8	9022	4.5	3294
		500	0.3%(0)	>1729k	1.4%(0)	>229k	0.4%(0)	>1441k
		1000	1.4%(0)	>893k	2.4%(0)	>168k	1.7%(0)	>832k
		3000	2.8%(0)	>215k	3.0%(0)	>61k	2.3%(0)	>257k

We observe from Table 3 that the performance of the branch-and-cut algorithm is improved by using strengthened big- M parameters. This is consistent with what has been shown in Table 1. From Table 1 we have seen that the root relaxation bound for the branch-and-cut algorithm is tighter than the MIP formulation (2.4). However, this advantage at the root node does not lead to an improvement in terms of the total computation time for solving these instances to optimality. It appears that branching in the Benders formulation is less effective than in the MIP formulation (2.4) and thus more nodes are explored. This motivates further study on effective ways to take advantage of the strong relaxation bound v_2^{LD} for solving CCSPs to optimality.

2.8.3 Performance on instances with binary x

We next consider the binary instances, i.e., with $S = \{0,1\}^n$. We compare the proposed dual bounds and also illustrate the effectiveness of the heuristic algorithm (Algorithm 1), the scenario decomposition algorithm (Algorithm 2) with and without scenario grouping, and the MIP formulation (2.4) with strengthened big- M coefficients. For the scenario

Table 4: Bound comparison for multi-dimensional binary knapsack instances

Instances	ϵ	N	$\underline{v}^C(M)$	z_1^{LP}	z_2^{NLP}	v^Q	v^{QG}	v^H
1-4-multi (20,10)	0.1	100	3.5%	3.5%	2.3%	1.6%	1.1%	0.0%
		500	3.8%	3.8%	2.6%	1.8%	1.5%	0.0%
		1000	3.8%	3.8%	2.7%	2.0%	1.8%	0.0%
		3000	3.8%	3.8%	2.7%	2.0%	1.7%	0.0%
	0.2	100	4.8%	4.7%	2.8%	2.3%	2.4%	0.0%
		500	3.9%	3.9%	2.2%	1.5%	1.6%	0.0%
		1000	4.4%	4.4%	2.7%	2.2%	2.2%	0.0%
		3000	4.4%	4.4%	2.7%	2.1%	2.0%	0.0%
1-6-multi (39,5)	0.1	100	3.2%	3.2%	2.7%	3.3%	2.2%	0.6%
		500	$\leq 3.9\%^*$	$\leq 3.9\%$	$\leq 3.3\%$	$\leq 3.9\%$	$\leq 3.1\%$	$\leq 0.3\%$
		1000	$\leq 4.0\%$	$\leq 4.0\%$	$\leq 3.5\%$	$\leq 4.1\%$	$\leq 3.4\%$	$\leq 1.6\%$
		3000	$\leq 4.0\%$	$\leq 4.0\%$	$\leq 3.5\%$	$\leq 4.2\%$	$\leq 3.3\%$	$\leq 2.7\%$
	0.2	100	3.9%	3.9%	2.9%	3.4%	3.0%	0.4%
		500	$\leq 4.2\%$	$\leq 4.2\%$	$\leq 3.0\%$	$\leq 3.3\%$	$\leq 3.6\%$	$\leq 1.1\%$
		1000	$\leq 4.4\%$	$\leq 4.3\%$	$\leq 3.3\%$	$\leq 3.8\%$	$\leq 4.0\%$	$\leq 2.3\%$
		3000	$\leq 4.5\%$	$\leq 4.5\%$	$\leq 3.5\%$	$\leq 4.0\%$	$\leq 3.8\%$	$\leq 3.2\%$

* A “ \leq ” indicates instances for which the optimal value is not known, and the associated number represents an upper bound on the true optimality gap.

decomposition algorithm, the lower bounds are obtained by v^Q . For the scenario grouping, the number of groups K is chosen as the smallest divisor of N that is larger than $\lceil \epsilon N \rceil$ and the scenarios are divided into K groups with the same size.

Table 4 summarizes the optimality gaps of $\underline{v}^C(M)$, z_1^{LP} , z_2^{NLP} , v^Q , v^{QG} and v^H , where v^{QG} denotes the results of the grouping quantile bound. For these instances, we report only gaps obtained with strengthened big- M coefficients. As we do not have the optimal solutions for most of the 1-6-multi instances, we use the best known upper bound or lower bounds to estimate the lower bound and heuristic gaps, respectively. Table 5 displays the time to compute these bounds. (Note that, the times for computing $\underline{v}^C(M)$, z_1^{LP} , and z_2^{NLP} are the times for $v^C(M)$, v_1^{LD} , and v_2^{LD} from Table 2 for the continuous x case, since these are equivalent.)

In Table 4, we see that the best Lagrangian dual bounds corresponding to continuous x still have at most 4% optimality gap, which demonstrates the effectiveness of these bounds. In addition, the quantile bound, which is obtained by solving binary IP subproblems, is somewhat stronger than any of the bounds $\underline{v}^C(M)$, z_1^{LP} , and z_2^{NLP} . On the other hand,

Table 5: Computational times for computing bounds for multi-dimensional binary knapsack instances

Instance	ϵ	N	$\underline{v}^C(M)$	z_1^{LP}	z_2^{NLP}	v^Q	v^{QG}	v^H
1-4-multi (20,10)	0.1	100	0.0	0.4	0.8	22.8	5.8	1.9
		500	0.0	3.4	6.2	136.0	31.7	12.6
		1000	0.2	11.8	21.5	282.8	62.5	47.6
		3000	2.0	84.4	151.9	770.4	134.8	285.8
	0.2	100	0.0	0.5	1.3	22.8	5.8	1.9
		500	0.1	3.6	12.5	136.0	31.7	11.7
		1000	0.2	13.7	39.0	282.8	62.5	48.1
		3000	2.3	93.5	292.0	770.4	134.8	340.6
1-6-multi (39,5)	0.1	100	0.0	0.5	1.2	31.8	6.7	2.7
		500	0.1	3.9	9.1	157.7	35.2	12.8
		1000	0.2	12.9	29.7	327.0	73.5	39.2
		3000	1.1	84.8	187.5	992.0	191.7	212.3
	0.2	100	0.0	0.7	2.2	31.8	6.7	2.4
		500	0.1	5.3	18.7	157.7	35.2	12.0
		1000	0.3	18.0	55.0	327.0	73.5	39.6
		3000	2.4	122.6	382.6	992.0	191.7	220.6

we see from Table 2 that the quantile bound v^Q takes longer to calculate. However, when we apply the quantile bound to the scenario grouping relaxation, the resulting bounds v^{QG} are comparable with the quantile bound obtained without grouping, but take much shorter time to compute (see Table 5).

We observe in Tables 4 and 5 that for the 1-4-multi instances the heuristic performs extremely well in terms of both quality (zero optimality gap) and solution time. For the larger instances, the optimality gaps are not exact since we cannot obtain the optimal objective values within the time limit, but we can see that the bounds obtained from the heuristic methods are still quite close to the optimal ones. Thus, the solution from the heuristic method could be treated as a good starting point for other algorithms.

Table 6 presents the results of solving these instances to optimality using the scenario decomposition algorithm (Algorithm 2) with and without scenario grouping, and the MIP formulation (2.4) with strengthened big- M coefficients. We find that the scenario grouping based decomposition method significantly outperforms the one without grouping in terms of computational time, for the instances solved within the time limit, and ending optimality

Table 6: Performance of scenario decomposition and MIP on the multi-dimensional binary knapsack instances

Instances			Scenario decomposition				MIP (2.4) With Big- M Str.		
			Grouping		Non-grouping				
Instance	ϵ	N	Time	Gap	Time	Gap	M Time	Tot. Time	Gap
1-4-multi (20,10)	0.1	100	23.3	0.0%(0)	111.4	0.0%(0)	2.4	3.4	0.0%(0)
		500	82.6	0.0%(0)	371.9	0.0%(0)	47.5	54.8	0.0%(0)
		1000	119.9	0.0%(0)	972.0	0.0%(0)	185.6	207.1	0.0%(0)
		3000	358.8	0.0%(0)	2253.1	0.0%(0)	1656.8	1837.4	0.0%(0)
	0.2	100	52.9	0.0%(0)	173.0	0.0%(0)	2.9	3.7	0.0%(0)
		500	122.9	0.0%(0)	357.1	0.0%(0)	47.4	56.8	0.0%(0)
		1000	232.3	0.0%(0)	1271.8	0.0%(0)	185.7	224.4	0.0%(0)
		3000	719.0	0.0%(0)	2017.3	0.0%(0)	1681.1	3281.4	1.4%(4)
1-6-multi (39,5)	0.1	100	3600.0	2.0%(5)	3600.0	3.6%(5)	1.0	4.7	0.0%(0)
		500	3600.0	2.4%(5)	3600.0	3.9%(5)	24.8	2619.9	0.1%(2)
		1000	3600.0	2.6%(5)	3600.0	4.0%(5)	98.8	3600.0	1.5%(5)
		3000	3600.0	3.0%(5)	3600.0	3.2%(5)	878.8	3600.0	2.7%(5)
	0.2	100	3600.0	2.6%(5)	3600.0	3.4%(5)	1.0	15.1	0.0%(0)
		500	3600.0	2.8%(5)	3600.0	3.2%(5)	24.7	3600.0	1.0%(5)
		1000	3600.0	3.6%(5)	1271.8	3.8%(5)	97.3	3600.0	2.3%(5)
		3000	3600.0	3.4%(5)	3600.0	3.7%(5)	878.9	3600.0	3.3%(5)

gaps, for the remaining instances. From Table 6, we further observe that when the number of scenarios is small (e.g., not larger than 1000), the MIP formulation (2.4) with strengthened big- M parameters gives the best performance among these three methods. However, when the number of scenarios is larger, the MIP formulation (2.4) could not close the optimality gap within the time limit, while the scenario decomposition method with grouping can still solve all of the 1-4-multi instances within 15 minutes. We also observe that neither method is able to solve the majority of the 1-6-multi instances.

CHAPTER III

ON QUANTILE CUTS AND THEIR CLOSURE FOR CHANCE CONSTRAINED OPTIMIZATION PROBLEMS

3.1 *Introduction*

A chance constrained problem (CCP) involves optimization over constraints (specified by stochastic data) which are required to be satisfied with a prescribed probability level. A generic formulation of CCP is

$$\min_x \left\{ c^\top x : x \in S, \mathbb{P}[\xi : x \in \mathcal{X}(\xi)] \geq 1 - \epsilon \right\}. \quad (3.1)$$

In the above formulation, S denotes a set of deterministic constraints, ξ denotes a random data vector, and $\mathcal{X}(\xi)$ denotes a system of stochastic constraints whose data is specified by the random vector ξ . The CCP (3.1) seeks a solution $x \in S$ that minimizes the cost $c^\top x$ and satisfies the stochastic constraints $\mathcal{X}(\xi)$ with probability at least $(1 - \epsilon)$ where $\epsilon \in (0, 1)$ is a prespecified risk level.

We consider a CCP with mixed integer convex constraints under finite distribution, i.e. we assume that

- $S = \{x \in \mathbb{R}^{n-\tau} \times \mathbb{Z}^\tau : G_0(x) \leq 0\}$ is a nonempty and compact deterministic mixed integer set defined by the convex mapping $G_0 : \mathbb{R}^n \rightarrow \mathbb{R}^{m_0}$;
- ξ is a random vector with a finite distribution supported on $\Xi = \{\xi^1, \dots, \xi^N\}$, where each ξ^i for $i \in \mathcal{N} := \{1, \dots, N\}$ corresponds to a scenario with a probability mass p_i ; and
- for a given scenario i , the vector ξ^i defines a nonempty and compact mixed convex integer constraint system $\mathcal{X}^i := \mathcal{X}(\xi^i) = \{x \in \mathbb{R}^{n-\tau} \times \mathbb{Z}^\tau : G_i(x) \leq 0\}$ defined by the convex mapping $G_i : \mathbb{R}^n \rightarrow \mathbb{R}^{m_i}$.

In this setting, the chance constraint in (3.1) corresponds to satisfying a subset $\mathcal{C} \subseteq [N]$

of the scenario constraints, i.e. $x \in \cap_{i \in \mathcal{C}} \mathcal{X}^i$, such that $\sum_{i \in \mathcal{C}} p_i \geq 1 - \epsilon$. Let

$$\mathcal{Z} := \left\{ \mathcal{C} \subseteq [N] : \sum_{i \in \mathcal{C}} p_i \geq 1 - \epsilon \right\}, \quad (3.2)$$

be the collection of all feasible subsets of scenarios. Then the feasible region of (3.1), denoted by X , can be written in the disjunctive normal form:

$$X = \bigcup_{\mathcal{C} \in \mathcal{Z}} \left[S \cap \mathcal{X}^i \right]. \quad (3.3)$$

We assume throughout that CCP is feasible, and hence X is nonempty. From the above disjunctive normal form it is clear that, even in the absence of integrality restrictions, i.e. $\tau = 0$, the set X is nonconvex, and not surprisingly CCP is strongly NP-hard [68, 82].

Since the sets \mathcal{X}^i for all $i \in [N]$ are compact we can introduce binary variables z_i for $i \in [N]$ and reformulate (3.1) as the mixed integer nonlinear program (MINLP):

$$\min_{x, z} \left\{ c^\top x : x \in S, G_i(x) \leq M_i(1 - z_i), \forall i \in \mathcal{N}, z \in Z \right\}, \quad (3.4)$$

where

$$Z := \left\{ z \in \{0, 1\}^N : \sum_{i \in [N]} p_i z_i \geq 1 - \epsilon \right\},$$

and $M_i \in \mathbb{R}^{m_i}$ for all $i \in [N]$ are suitable big-M coefficients. Since the continuous relaxation of (3.4) is typically very weak, there has been a great deal of work in deriving strong valid inequalities for this MINLP. One popular approach is to derive a relaxation of (3.4) in the form of the well-studied mixing set [45] and add the corresponding mixing inequalities [1, 56, 66, 68, 133].

We consider a family of valid inequalities for the nonconvex feasible region X of the CCP (3.1) in the original x -space, rather than those for the MINLP formulation (3.4) in the (x, z) -space. These valid inequalities known as *quantile cuts* are obtained as follows. We first optimize a linear function $\alpha^\top x$ over each scenario constraint, and record the optimal values $\beta_i^\alpha = \min\{\alpha^\top x : x \in S \cap \mathcal{X}^i\}$ for $i \in [N]$. This approach and resulting β_i^α values was used in [66] to derive a mixing set relaxation for (3.4). Notice that each β_i^α has the associated probability p_i . Next we compute the $(1 - \epsilon)$ -quantile of $\{\beta_i^\alpha\}_{i \in [N]}$ based on these probabilities – denote this by β_q^α . The quantile cut is then given by $\alpha^\top x \geq \beta_q^\alpha$. Such

inequalities were studied in [82] where it is shown that a single quantile cut represents the projection of the convex hull of a mixing set relaxation of (3.4) in the (x, z) -space onto the x space. Quantile cuts have been used in computational studies of chance constrained problems with good results [4, 82, 101].

In this chapter we undertake a theoretical study of quantile cuts. In particular we study properties of the quantile closure, i.e. the intersection of all quantile cuts. Quantile cuts represent an infinite family of inequalities - one for each α vector - and so a finite description of the quantile closure and separation over it are important questions. By replacing the deterministic constraint system S in (3.1) with the (first) quantile closure we obtain a stronger formulation for which we can apply another round of quantile cuts and derive the second quantile closure and so on. We investigate how the sequence of sets produced by such successive quantile closure operations relates to the convex hull of the feasible region of (3.1). The main results of this chapter are summarized below.

1. We show that the quantile closure has a finite description in conjunctive normal form. An important corollary of this result is that for a mixed integer linear CCP, the quantile closure is polyhedral.
2. We prove that the sequence of sets obtained by successive quantile closure operations converges to the convex hull of X (i.e., $\text{conv}(X)$) with respect to the Hausdorff metric. In the pure integer linear setting, i.e. $\tau = n$, there exists a finite number of rounds of the quantile closure that recovers $\text{conv}(X)$.
3. We study an approximation of the quantile closure given by a restricted choice of α , and show finite convergence for mixed integer linear CCPs. We also study the approximation error for covering problems.
4. We study a generalization and strengthening quantile cuts and their closure obtained by grouping scenarios together.
5. We show that separation over the first quantile closure is NP-hard, and propose a

heuristic separation algorithm. We present a computational study to show the effectiveness of heuristically separated quantile cuts.

The remainder of this chapter is organized as follows. In Section 3.2 we discuss the connection of quantile cuts for $\text{conv}(X)$ to the mixing set inequalities for the MINLP (3.4). In Section 3.3 we establish the conjunctive normal form of the quantile closures. In Section 3.4 we study convergence properties of successive quantile closures. An approximate quantile closure is studied in Section 3.5 followed by a generalization by grouping scenarios together in Section 3.6. In Section 3.7, we prove complexity of separation over the first quantile closure and propose a heuristic separation method. Numerical studies in Section 3.8 show that these quantile cuts indeed help reduce the root gap as well as the solution time.

3.2 Quantile cuts and Mixing inequalities

We first formally define the quantile cut for the CCP (3.1). Recall that S is a set of deterministic constraints, \mathcal{X}^i are constraints associated with scenario $i \in [N]$, \mathcal{Z} defined in (3.2) is the collection of all feasible scenario sets, and X given by (3.3) is the set of feasible solutions of (3.1).

Definition 1. Given $\alpha \in \mathbb{R}^n$ let $\{\beta_i^\alpha(S)\}_{i \in \mathcal{N}}$ be the optimal values of

$$\beta_i^\alpha(S) = \min \left\{ \alpha^\top x : x \in S \cap \mathcal{X}^i \right\} \quad \forall i \in \mathcal{N}. \quad (3.5)$$

The quantile $\beta_q^\alpha(S)$ is given by

$$\beta_q^\alpha(S) := \min_{\mathcal{C} \in \mathcal{Z}} \max_{i \in \mathcal{C}} \beta_i^\alpha(S) \quad (3.6)$$

and the associated “quantile cut” is

$$\alpha^\top x \geq \beta_q^\alpha(S). \quad (3.7)$$

Note that the above definition depends on S so as to allow for successive applications with changing S . Since S and \mathcal{X}^i are compact we have that $\beta_i^\alpha(S) \in (-\infty, +\infty]$ where the value of $+\infty$ is taken when the problem (3.5) is infeasible. When $p_i = \frac{1}{N}$ for all $i \in \mathcal{N}$, we have that $\beta_q^\alpha(S)$ is the $(\lfloor \epsilon N \rfloor + 1)$ st largest value among $\{\beta_i^\alpha(S)\}_{i \in \mathcal{N}}$.

From the definition above and the disjunctive normal form (3.3) of X it should be clear that the quantile cut (3.7) is valid for $\text{conv}(X)$. We next reveal the connection between quantile cuts and mixing inequalities for CCP, which also establishes the validity of these cuts.

A mixing set [45] is a mixed-integer set of the form

$$P = \{(v, z) \in \mathbb{R}_+ \times \{0, 1\}^s : v + h_i z_i \geq h_i \ i = 1, \dots, s\} \quad (3.8)$$

with $h_1 \geq \dots \geq h_s$. The following exponential family of mixing inequalities are valid for P

$$v + \sum_{j=1}^l (h_{t_j} - h_{t_{j+1}}) z_{t_j} \geq h_{t_1} \quad \forall T = \{t_1, \dots, t_l\} \subseteq \{1, \dots, s\}, \quad (3.9)$$

where $h_{t_1} \geq \dots \geq h_{t_l}, h_{t_{l+1}} = 0$. These inequalities are facet defining for P when $t_1 = 1$ and are sufficient to describe the convex hull of P (see [7, 45]).

Using the β -values as defined in (3.5) Luedtke [66] constructed the following mixing set relaxation of CCP (3.4). Note that in [66] each scenario \mathcal{X}^i is a polyhedron, however the construction directly extends to the MINLP formulation (3.4).

$$Y^\alpha = \left\{ (x, z) \in \mathbb{R}^n \times \{0, 1\}^N : \alpha^\top x + (\beta_i^\alpha(S) - \beta_q^\alpha(S))(1 - z_i) \geq \beta_i^\alpha(S), \right. \\ \left. i \in \mathcal{B}_q^\alpha, z \in Z \right\}. \quad (3.10)$$

where $\mathcal{B}_q^\alpha := \{i \in [N] : \beta_i^\alpha(S) \geq \beta_q^\alpha(S)\}$ is a subset of scenarios each of whose β -value is at least as large as the quantile $\beta_q^\alpha(S)$.

Proposition 14. (*Theorem 1, [66]*) *For any α , the system Y^α is a relaxation of the feasible region of the MINLP (3.4), and hence $X \subseteq \text{Proj}_x(Y^\alpha)$, where $\text{Proj}_x(\cdot)$ denotes the projection of a set onto the x -space.*

Note that Y^α is a mixing system with a knapsack side constraint defined by Z , thus the mixing inequalities of the form (3.9) are valid. These inequalities were used within a branch and cut scheme for solving the MINLP (3.4) in [66]. Note that the mixing inequalities are in the (x, z) -space while the quantile cuts are in the original x -space. The next result shows that a single quantile cut in the x -space captures the effect of the entire exponential family of mixing inequalities.

Proposition 15. (*Proposition 5, [82]*) For any α ,

$$\text{Proj}_x(\text{conv}(Y^\alpha)) = \{x \in \mathbb{R}^n : \alpha^\top x \geq \beta_q^\alpha(S)\}.$$

Inspired by the above result we investigate, in the remainder of the chapter, the strength of the quantile closure, i.e. the intersection of all quantile cuts.

3.3 Quantile closure

In this section we define quantile closures and establish their finite characterizations.

Definition 2. The first quantile closure of S is defined as

$$S^1 := \bigcap_{\alpha \in \mathbb{R}^n} \left\{ x \in \mathbb{R}^n : \alpha^\top x \geq \beta_q^\alpha(S) \right\}.$$

Inductively, we define r th round quantile closure S^r as

$$S^r := \bigcap_{\alpha \in \mathbb{R}^n} \left\{ x \in \mathbb{R}^n : \alpha^\top x \geq \beta_q^\alpha(S^{r-1}) \right\} \quad r \geq 2.$$

Next we characterize $\text{conv}(X)$ and S^1 in conjunctive normal form. Let us begin with the following definition.

Definition 3. A set $\mathbf{g} \subseteq \mathcal{N}$ is a “partial covering subset” if it intersects with all of feasible scenario subsets in \mathcal{Z} , i.e., for any $\hat{\mathcal{C}} \in \mathcal{Z}$, we have $\mathbf{g} \cap \hat{\mathcal{C}} \neq \emptyset$. Also, a set \mathbf{g} is a “minimal” partial covering subset if there does not exist another partial covering subset $\mathbf{g}' \subseteq \mathcal{N}$ such that $\mathbf{g}' \subsetneq \mathbf{g}$. We let \mathcal{G} denote the collection of all minimal partial covering subsets.

Note that when $p_i = \frac{1}{N}$ for all $i \in \mathcal{N}$, then the collection of minimal partial covering subsets is $\mathcal{G} = \{\mathbf{g} \subseteq \mathcal{N} : |\mathbf{g}| = \lfloor \epsilon N \rfloor + 1\}$.

Proposition 16.

$$X = \bigcap_{\mathbf{g} \in \mathcal{G}} \left[\bigcup_{i \in \mathbf{g}} (S \cap \mathcal{X}^i) \right]. \quad (3.11)$$

Proof. Define $X' = \bigcap_{\mathbf{g} \in \mathcal{G}} \left[\bigcup_{i \in \mathbf{g}} (S \cap \mathcal{X}^i) \right]$. We need to show that $X = X'$.

Let $x \in X$. Then, there exists a feasible subset $\mathcal{C} \in \mathcal{Z}$ such that $x \in \bigcap_{i \in \mathcal{C}} (S \cap \mathcal{X}^i)$. For an arbitrary minimal partial covering $\mathbf{g} \in \mathcal{G}$, we must have $x \in \bigcup_{i \in \mathbf{g}} (S \cap \mathcal{X}^i)$ since from Definition 3, \mathbf{g} intersects with all feasible subsets (i.e., $\mathbf{g} \cap \mathcal{C} \neq \emptyset$). Thus, $X \subseteq X'$.

Suppose that there exists an $x' \in X'$ such that $x' \notin X$. Define a subset $\mathcal{C}' := \{i \in \mathcal{N} : x' \in S \cap \mathcal{X}^i\} \notin \mathcal{Z}$. Let \mathbf{g}' be the complement of \mathcal{C}' , i.e., $\mathbf{g}' = \mathcal{N} \setminus \mathcal{C}'$. We claim that for all $\mathcal{C} \in \mathcal{Z}$, we have $\mathbf{g}' \cap \mathcal{C} \neq \emptyset$. Suppose not, then there must exist a $\widehat{\mathcal{C}} \in \mathcal{Z}$ such that $\mathbf{g}' \cap \widehat{\mathcal{C}} = \emptyset$. This implies that $\widehat{\mathcal{C}} \subseteq \mathcal{C}'$, and thus

$$x' \in \bigcap_{i \in \mathcal{C}'} (S \cap \mathcal{X}^i) \subseteq \bigcap_{i \in \widehat{\mathcal{C}}} (S \cap \mathcal{X}^i) \subseteq X,$$

which contradicts $x' \notin X$. Hence, \mathbf{g}' is a partial covering subset of \mathcal{N} , and thus $x' \in X' \subseteq \bigcup_{i \in \mathbf{g}'} (S \cap \mathcal{X}^i)$; this contradicts the definition of $\mathbf{g}' = \mathcal{N} \setminus \mathcal{C}'$. \square \square

Next we provide a conjunctive normal form for S^1 . We will need the following preliminary observations.

Lemma 1. *The set $\mathcal{B}_q^\alpha = \{i \in [N] : \beta_i^\alpha(S) \geq \beta_q^\alpha(S)\}$ is a partial covering subset.*

Proof. From the definition of $\beta_q^\alpha(S)$, for any subset $\widehat{\mathcal{C}} \in \mathcal{Z}$, there must exist an $i_0 \in \widehat{\mathcal{C}}$ such that $\beta_{i_0}^\alpha(S) \geq \beta_q^\alpha(S)$. Thus \mathcal{B}_q^α is a partial covering subset. \square \square

Lemma 2. *There exist a $\mathbf{g} \in \mathcal{G}$ such that $\beta_q^\alpha(S) = \min_{i \in \mathbf{g}} \beta_i^\alpha(S)$.*

Proof. By the definition of $\beta_q^\alpha(S)$, there exists a $\widehat{\mathcal{C}} \in \mathcal{Z}$ with $\beta_q^\alpha(S) = \max_{\widehat{\mathcal{C}}} \beta_i^\alpha(S) \geq \beta_j^\alpha(S)$ for all $j \in \widehat{\mathcal{C}}$. From Definition 3, for each $\widehat{\mathbf{g}} \in \mathcal{G}$, we have $\widehat{\mathbf{g}} \cap \widehat{\mathcal{C}} \neq \emptyset$. Hence, $\beta_q^\alpha(S)$ must be no smaller than the smallest value in set $\{\beta_i^\alpha(S)\}_{i \in \widehat{\mathbf{g}}}$; i.e.,

$$\beta_q^\alpha(S) \geq \min_{i \in \widehat{\mathbf{g}} \cap \widehat{\mathcal{C}}} \beta_i^\alpha(S) \geq \min_{i \in \widehat{\mathbf{g}}} \beta_i^\alpha(S).$$

From Lemma 1, \mathcal{B}_q^α is a partial covering subset. Now let \mathbf{g} be a minimal partial covering subset such that $\mathbf{g} \subseteq \mathcal{B}_q^\alpha$. Thus,

$$\min_{i \in \mathbf{g}} \beta_i^\alpha(S) \geq \min_{i \in \mathcal{B}_q^\alpha} \beta_i^\alpha(S) \geq \beta_q^\alpha(S).$$

\square

\square

Theorem 9.

$$S^1 = \bigcap_{\mathbf{g} \in \mathcal{G}} \text{conv} \left[\bigcup_{i \in \mathbf{g}} (S \cap \mathcal{X}^i) \right], \quad (3.12)$$

and for each $r \geq 2, r \in \mathbb{Z}_{++}$,

$$S^r = \bigcap_{\mathbf{g} \in \mathcal{G}} \text{conv} \left[\bigcup_{i \in \mathbf{g}} (S^{r-1} \cap \mathcal{X}^i) \right]$$

where $S^0 = S$.

Proof. (i) We first prove (3.12).

Let $W^{\mathbf{g}} := \text{conv} \left[\bigcup_{i \in \mathbf{g}} (S \cap \mathcal{X}^i) \right]$, $W := \bigcap_{\mathbf{g} \in \mathcal{G}} W^{\mathbf{g}}$. We need to show that $S^1 = W$.

$[S^1 \subseteq W]$ Consider $\mathbf{g} \in \mathcal{G}$, and take any valid inequality $\alpha^\top x \geq \beta$ for $W^{\mathbf{g}}$. Let $\widehat{C} \in \mathcal{Z}$ such that $\beta_q^\alpha(S) = \max_{i \in \widehat{C}} \beta_i^\alpha(S)$. Since $\mathbf{g} \cap \widehat{C} \neq \emptyset$ by Definition 3, hence

$$\beta_q^\alpha(S) \geq \min_{i \in \mathbf{g} \cap \widehat{C}} \beta_i^\alpha(S) \geq \min_{i \in \mathbf{g}} \beta_i^\alpha(S) \geq \beta.$$

Thus, $\alpha^\top x \geq \beta$ is a valid inequality of S^1 . This holds for any valid inequality of $W^{\mathbf{g}}$, we have that $S^1 \subseteq W^{\mathbf{g}}$. Since \mathbf{g} was arbitrary, it follows that $S^1 \subseteq W^{\mathbf{g}}$ for all $\mathbf{g} \in \mathcal{G}$; i.e., $S^1 \subseteq W$.

$[S^1 \supseteq W]$ For any given α , from Lemma 2, there exist a $\mathbf{g} \in \mathcal{G}$ such that $\beta_q^\alpha(S) = \min_{i \in \mathbf{g}} \beta_i^\alpha(S)$. Clearly, $\alpha^\top x \geq \beta_q^\alpha(S)$ is a valid inequality for $W^{\mathbf{g}}$; and so it is valid for W . Thus, $S^1 \supseteq W$.

(ii) When $r \geq 2$, the statement follows directly from (3.12) by replacing S with S^{r-1} .

□

□

Next we show that the conjunctive normal form (3.12) of S^1 which is independent of α implies the polyhedrality of the quantile closures when $G_0(\cdot), \{G_i(\cdot)\}_{i \in \mathcal{N}}$ are rational affine mappings.

Corollary 2. *For each $r \in \mathbb{Z}_{++}$, if $G_0(\cdot), \{G_i(\cdot)\}_{i \in \mathcal{N}}$ are rational affine functions, then S^r is a polytope.*

Proof. By the fundamental theorem of mixed integer program [72] and the fact that the convex hull of union of compact sets is equivalent to the convex hull of the union of convex hulls of compact sets, it follows that for any $\mathbf{g} \in \mathcal{G}$,

$$\text{conv} \left[\bigcup_{i \in \mathbf{g}} (S \cap \mathcal{X}^i) \right] = \text{conv} \left[\bigcup_{i \in \mathbf{g}} \text{conv} (S \cap \mathcal{X}^i) \right]$$

and is a polytope. Since \mathcal{G} is a finite set, it follows from Theorem 9 that S^1 is a polytope.

By induction, suppose S^r is a polytope for $r \leq t$. Now let $r = t + 1$, by (i), we have

$$S^r = \bigcap_{\mathbf{g} \in \mathcal{G}} \text{conv} \left[\bigcup_{i \in \mathbf{g}} (S^{r-1} \cap \mathcal{X}^i) \right]$$

and S^{r-1} is a polytope, hence S^r is a polytope. □ □

3.4 Convergence of quantile closures

In this section, we investigate convergence of successive rounds of quantile closure operations. Our convergence notions are with respect to the Hausdorff distance [92]. For two closed convex sets $K_1, K_2 \in \mathbb{R}^n$, the Hausdorff distance $d_H(K_1, K_2)$ is defined as

$$\begin{aligned} d_H(K_1, K_2) &:= \min \{ \delta : K_1 \subseteq K_2 + B(0, \delta), K_2 \subseteq K_1 + B(0, \delta) \} \\ &:= \max \left\{ \sup_{x \in K_1} \inf_{y \in K_2} \|x - y\|_2, \sup_{x \in K_2} \inf_{y \in K_1} \|x - y\|_2 \right\}, \end{aligned}$$

where $B(0, \delta)$ denotes the ball centered at origin with radius δ . We will need the following fact on the limit of a set sequence.

Lemma 3. (Proposition 2, [91]) *Let $\{R^r\}$ be a sequence of nonempty closed convex sets such that $R^{r+1} \subseteq R^r$ for all r . Then R^r converges to $\bar{R} := \lim_{r \rightarrow \infty} R^r = \bigcap_{r=1}^{\infty} R^r$ with respect to the Hausdorff distance, and \bar{R} is also a closed convex set.*

The following lemma reveals the convergence properties of a sequence of sets produced by successive quantile closure operations.

Lemma 4. *Let $\{S^r\}$ be a sequence of quantile closures. Then*

(i) *there exists a $\bar{S} := \lim_{r \rightarrow \infty} S^r$;*

(ii) for each $\mathbf{g} \in \mathcal{G}$, we have

$$\text{conv} \left[\bigcup_{i \in \mathbf{g}} \text{conv} (\bar{S} \cap \mathcal{X}^i) \right] = \bar{S}. \quad (3.13)$$

Proof. (i) This directly follows from Lemma 3 since $\{S^r\}$ is an inclusion-wise monotone sequence of convex sets.

(ii) Let \bar{S}^1 be the quantile closure operation applied to set \bar{S} . Since $\bar{S} = \bar{S}^1$ by the limiting operation, we have that

$$\bar{S} = \bar{S}^1 = \bigcap_{\mathbf{g} \in \mathcal{G}} \text{conv} \left[\bigcup_{i \in \mathbf{g}} (\bar{S} \cap \mathcal{X}^i) \right] = \bigcap_{\mathbf{g} \in \mathcal{G}} \text{conv} \left[\bigcup_{i \in \mathbf{g}} \text{conv} (\bar{S} \cap \mathcal{X}^i) \right],$$

where the second equality is due to Theorem 9 and the third equality follows from the fact that the convex hull of union of compact sets is equivalent to the convex hull of union of convex hull of compact sets. Since $\text{conv} \left[\bigcup_{i \in \mathbf{g}} (\bar{S} \cap \mathcal{X}^i) \right] \subseteq \bar{S}$ for all $\mathbf{g} \in \mathcal{G}$, we have that (3.13) holds.

□

□

Now, we are ready to prove the convergence of the quantile closure procedure to the convex hull of X .

Theorem 10. *The set sequence $\{S^r\}$ converges to $\text{conv}(X)$ with respect to the Hausdorff distance; i.e., $\bar{S} = \lim_{r \rightarrow \infty} S^r = \text{conv}(X)$.*

Proof. From Lemma 3, we know that there exists an $\bar{S} = \lim_{r \rightarrow \infty} S^r$. Since $\text{conv}(X) \subseteq S^r$ for all r , it follows that $\text{conv}(X) \subseteq \bar{S}$. Thus, we only need to show that $\text{conv}(X) \supseteq \bar{S}$. It suffices to show that any extreme point of the convex set \bar{S} belongs to X which will complete the proof.

Consider an extreme point \bar{x} of \bar{S} . By the identity (3.13) in Lemma 4 and the facts that $\bar{S} \subseteq \text{conv}(S)$ and any extreme point of the convex hull of union of compact sets comes from at least one of the compact sets, it follows that there exists an $i_{\mathbf{g}} \in \mathbf{g}$ such that $\bar{x} \in \bar{S} \cap \mathcal{X}^{i_{\mathbf{g}}} \subseteq S \cap \mathcal{X}^{i_{\mathbf{g}}}$ for each $\mathbf{g} \in \mathcal{G}$. Let $\bar{\mathcal{C}} := \{i \in \mathcal{N} : \bar{x} \in S \cap \mathcal{X}^i\}$. We make the following claim.

Claim: $\bar{\mathcal{C}} \in \mathcal{Z}$.

Proof. Suppose not. Let $\bar{\mathbf{g}}$ be the complement of $\bar{\mathcal{C}}$, i.e., $\bar{\mathbf{g}} = \mathcal{N} \setminus \bar{\mathcal{C}}$. First of all, note that we have $\bar{\mathbf{g}} \cap \mathcal{C} \neq \emptyset$ for all $\mathcal{C} \in \mathcal{Z}$. Otherwise, there must exist a $\hat{\mathcal{C}} \in \mathcal{Z}$ and $\bar{\mathbf{g}} \cap \hat{\mathcal{C}} = \emptyset$, which implies that $\hat{\mathcal{C}} \subseteq \bar{\mathcal{C}}$, a contradiction that $\bar{\mathcal{C}} \notin \mathcal{Z}$. Hence, $\bar{\mathbf{g}}$ is a partial covering subset of \mathcal{N} . Let $\hat{\mathbf{g}}$ be a minimal partial covering subset such that $\hat{\mathbf{g}} \subseteq \bar{\mathbf{g}}$. Since we know that $\bar{x} \in S \cap X^{i_{\hat{\mathbf{g}}}}$ for some $i_{\hat{\mathbf{g}}} \in \hat{\mathbf{g}}$ (i.e., $i_{\hat{\mathbf{g}}} \in \bar{\mathcal{C}}$), we have a contradiction that $\hat{\mathbf{g}} \cap \bar{\mathcal{C}} = \emptyset$. \diamond \square

It then follows that $\bar{x} \in \bigcap_{i \in \bar{\mathcal{C}}} S \cap \mathcal{X}^i \subseteq X$. This completes the proof. \square \square

Next we show that in the pure integer setting the convex hull of X can be obtained after a finite number of quantile closure operations.

Theorem 11. *Suppose that $S \cap \mathcal{X}^i \subseteq \mathbb{Z}^n$ for all $i \in \mathcal{N}$ (i.e., $\tau = n$), then there exists a finite \bar{r} such that*

$$\bar{S} = S^{\bar{r}} = \text{conv}(X).$$

Proof. From Theorem 10, we know that $\bar{S} = \text{conv}(X)$. Now we only need to show the finite convergence.

Claim 1: If $\text{conv}(S^r \cap \mathbb{Z}^n) \neq \text{conv}(X)$, then there exists a $\delta > 0$ (irrespective of r) such that $d_H(S^r, \text{conv}(X)) \geq \delta$.

Proof. First of all, we know that all of the extreme points of $\text{conv}(S \bigcap_{i \in \mathcal{C}} \mathcal{X}^i)$ is integral. By the fact that any extreme point of the convex hull of union of compact sets comes from at least one of the compact sets, $\text{conv}(X)$ is integral; i.e., all of the extreme points of $\text{conv}(X)$ is integral.

If $\text{conv}(S^r \cap \mathbb{Z}^n) \neq \text{conv}(X)$, then there must exist a vector $\bar{x}^0 \in \mathbb{Z}^n \setminus \text{conv}(X)$ such that $\bar{x}_0 \in S^r$, but $\bar{x}_0 \notin \text{conv}(X)$. Thus, $d_H(S^r, \text{conv}(X))$ is lower bounded by the Hausdorff distance between $\mathbb{Z}^n \setminus \text{conv}(X)$ and $\text{conv}(X)$, which is at least

$$\sup_{y \in \text{conv}(X)} \inf_{x \in \mathbb{Z}^n \setminus \text{conv}(X)} \|x - y\|_2$$

Note that the $\mathbf{ext}(\text{conv}(X)) \subseteq \text{conv}(X)$, where $\mathbf{ext}(Y)$ denotes the set of extreme points of closed convex set Y . Thus, $d_H(S^r, \text{conv}(X))$ is no smaller than

$$\sup_{\mathbf{ext}(\text{conv}(X))} \inf_{x \in \mathbb{Z}^n \setminus \text{conv}(X)} \|x - y\|_2$$

which is clearly greater than or equal to $\delta = 1$.

◇

□

It then follows that there must exist a $\bar{r} \in \mathbb{Z}_{++}$ such that $\text{conv}(S^{\bar{r}-1} \cap \mathbb{Z}^n) = \text{conv}(X)$; otherwise, by Claim 1, $d_H(S^r, \text{conv}(X)) \geq \delta$ for all r , contradicting the fact that $\lim_{r \rightarrow \infty} S^r = \text{conv}(X)$. Since $\text{conv}(S^{\bar{r}-1} \cap \mathbb{Z}^n) = \text{conv}(X)$, then by Theorem 9, we have $S^{\bar{r}} = \text{conv}(X) := \bar{S}$. □

We close this section with two examples. The first shows the necessity of the compactness assumption for the convergence of the quantile closure to the convex hull, and second shows the necessity of the pure integer setting for finite convergence.

Example 2. Let $S = \mathbb{R}^2$, $\mathcal{X}^1 = \{x \in \mathbb{R}^2 : 0 \leq x_1 \leq 2, x_2 = 0\}$, $\mathcal{X}^2 = \{x \in \mathbb{R}^2 : x_1 = 0, x_2 \geq 0\}$, $\mathcal{X}^3 = \{x \in \mathbb{R}^2 : x_1 = 2, x_2 \geq 0\}$, $\epsilon = \frac{1}{3}$, $p_i = \frac{1}{3}$, $i = 1, 2, 3$ (see Figure 2 for an illustration). Since each feasible set contains at least two scenarios, by (3.3), we have $\text{conv}(X) = \{x \in \mathbb{R}^2 : 0 \leq x_1 \leq 2, x_2 = 0\}$. As there are exactly two scenarios in each minimal partial covering subset, according to Theorem 9, we have $S^1 = \dots = S^r = \dots = \bar{S} = \{x \in \mathbb{R}^2 : 0 \leq x_1 \leq 2, x_2 \geq 0\}$. Hence, in this example, the scenario constraints do not define bounded feasible regions, and the quantile closures do not converge to the convex hull of the feasible region X ; i.e., $\bar{S} \neq \text{conv}(X)$. ◇

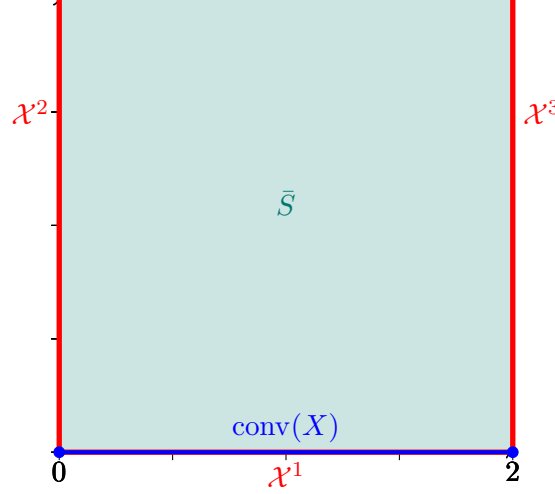


Figure 2: Illustration of Example 2

Example 3. Suppose $S = [0, 2]^2$, $\mathcal{X}^1 = \{x \in \mathbb{R}_+^2 : 2x_1 + 0.5x_2 \geq 1\}$, $\mathcal{X}^2 = \{x \in \mathbb{R}_+^2 : 0.5x_1 + 2x_2 \geq 1\}$, $\mathcal{X}^3 = \{x \in \mathbb{R}_+^2 : x_1 + x_2 \geq 1\}$, $\epsilon = \frac{1}{3}$, $p_i = \frac{1}{3}$, $i = 1, 2, 3$ (see Figure 3 for an illustration). Since each feasible set contains at least two scenarios, by (3.3), we have

$$\text{conv}(X) = \text{conv}\{(1, 0), (0.4, 0.4), (0, 1), (0, 2), (2, 0), (2, 2)\},$$

which contains the set \mathcal{X}^3 . By induction, we can show that

$$S^r = \text{conv}\{(1, 0), (w_r, w_r), (0, 1), (0, 2), (2, 0), (2, 2)\},$$

where $0 < w_r < 0.4$ for all $r \in \mathbb{Z}_{++}$; i.e., $S^r \neq \text{conv}(X)$ whenever $r < \infty$.

Indeed, when $r = 1$, as there are exactly two scenarios in each minimal partial covering subset, according to (3.12), we have

$$S^1 = \text{conv}\{(1, 0), (1/3, 1/3), (0, 1), (0, 2), (2, 0), (2, 2)\},$$

where $w_1 = 1/3 \in (0, 0.4)$. Suppose for $\gamma = r \geq 1$, the hypothesis holds; i.e.,

$$S^r = \text{conv}\{(1, 0), (w_r, w_r), (0, 1), (0, 2), (2, 0), (2, 2)\},$$

where $0 < w_r < 0.4$. Now let $\gamma = r + 1$, then by Theorem 9, we have

$$S^{r+1} = \text{conv}\{(1, 0), (w_{r+1}, w_{r+1}), (0, 1), (0, 2), (2, 0), (2, 2)\},$$

where $w_{r+1} = 0.3 + 1/(30 - 50w_r) \in (0, 0.4)$. \diamond

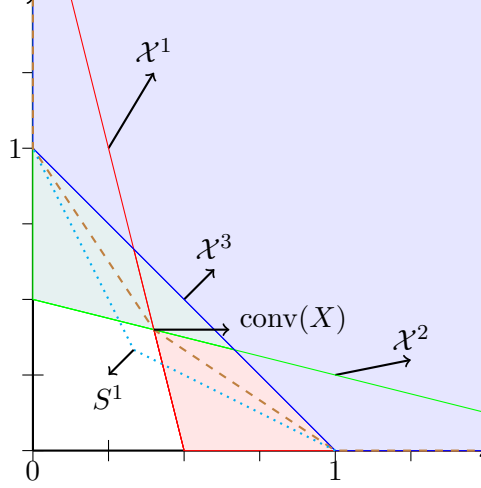


Figure 3: Illustration of Example 3

3.5 Approximate quantile closure in the polyhedral setting

In this section, we assume that $G_0(\cdot), \{G_i(\cdot)\}_{i \in \mathcal{N}}$ are rational affine mappings, in particular, $G_0(x) := d - Dx$ and $G_i(x) := b^i - A^i x$ for each $i \in \mathcal{N}$ where $D \in \mathbb{Z}^{m_0 \times n}, d \in \mathbb{Z}^{m_0}$ and $A^i \in \mathbb{Z}^{m_i \times n}, b^i \in \mathbb{Z}^{m_i}$. We consider quantile cuts derived by restricting the choice of α to the rows of D and $\{A^i\}_{i \in \mathcal{N}}$. Such cuts we considered in numerical studies in [82, 101].

3.5.1 The approximate scheme

Given a valid inequality $\alpha^\top x \geq \beta$ of $\text{conv}(X)$ with α chosen from the rows of D and $\{A^i\}_{i \in \mathcal{N}}$, we first show that the size of β cannot be arbitrarily large.

Proposition 17. *Suppose $\alpha^\top x \geq \beta$ is a valid inequality of $\text{conv}(X)$ with α chosen from $\{D_t\}_{t \in [m_0]} \cup \{A_t^i\}_{t \in [m_i], i \in \mathcal{N}}$. Then $\alpha^\top x \geq \bar{\beta}$ is also a valid inequality for $\text{conv}(X)$, where*

$$\bar{\beta} := \min_{\bar{q}_1, \bar{q}_2 \in \mathbb{Z}} \left\{ \frac{\bar{q}_1}{\bar{q}_2} : \frac{\bar{q}_1}{\bar{q}_2} \geq \beta, |\bar{q}_2| \leq 2^{\phi+\psi}, |\bar{q}_2| \leq 2^\phi, \gcd(|\bar{q}_1|, |\bar{q}_2|) \equiv 1 \right\}, \quad (3.14)$$

where ϕ, ψ are two positive integer numbers.

Proof. (i) First of all, from (3.3), we note that for an arbitrary extreme point \hat{x} of $\text{conv}(X)$, there exists a collection $\mathcal{C} \in \mathcal{Z}$ such that \hat{x} is also an extreme point of $\text{conv}(S \cap_{i \in \mathcal{C}} \mathcal{X}^i)$. Since $\text{conv}(S), \{\text{conv}(\mathcal{X}^i)\}_{i \in \mathcal{N}}$ are nonempty polytopes, and are defined by rational data, thus all the extreme points of $\text{conv}(S \cap_{i \in \mathcal{C}} \mathcal{X}^i)$ are rational, and

there exists a positive integer ϕ such that the encoding length of each extreme point of $\text{conv}(X)$ is bounded by ϕ . Here, the encoding length of a rational vector $x \in \mathcal{Q}^n$, where $x_j = \frac{q_{1j}}{q_{2j}}$, $q_{1j}, q_{2j} \in \mathbb{Z}$ with $|q_{1j}|, |q_{2j}|$ relatively prime for each $j \in [n]$, is defined as

$$n + \sum_{j \in [n]} [1 + \lceil \log_2(|q_{1j}| + 1) \rceil + \lceil \log_2(|q_{2j}| + 1) \rceil].$$

(ii) Next, suppose that the largest encoding length of each row of matrices $D, \{A^i\}_{i \in \mathcal{N}}$ is at most ψ . Then for an $\alpha \in \{D_{t\cdot}\}_{t \in [m_0]} \cup \{A_{t\cdot}^i\}_{t \in [m_i], i \in \mathcal{N}}$, optimizing $\alpha^\top x$ over $\text{conv}(X)$ is achieved by an extreme point \hat{x} . Let $\alpha^\top \hat{x} := \frac{\hat{q}_1}{\hat{q}_2}$ where $\hat{q}_1, \hat{q}_2 \in \mathbb{Z}$ with $|\hat{q}_1|, |\hat{q}_2|$ relatively prime. Note that the encoding length of $|\hat{q}_1| \leq 2^{\phi+\psi}, |\hat{q}_2| \leq 2^\phi$ because the encoding length of each extreme point of $\text{conv}(X)$ is bounded by ϕ and $\alpha \in \mathbb{Z}^n$ is of encoding length at most ψ . Thus, for any valid inequality $\alpha^\top x \geq \beta$, $\frac{\hat{q}_1}{\hat{q}_2}$ is a feasible solution to (3.14). Thus, $\alpha^\top x \geq \frac{\hat{q}_1}{\hat{q}_2} \geq \bar{\beta}$ is also a valid inequality to $\text{conv}(X)$. \square

\square

We now formally define the rounds of the approximate quantile closure below.

Definition 4. *The r th round of approximate quantile closure of S is defined as*

$$\hat{S}^r := \bigcap_{\alpha \in \{D_{t\cdot}\}_{t \in [m_0]} \cup \{A_{t\cdot}^i\}_{t \in [m_i], i \in \mathcal{N}}} \left\{ x \in \mathbb{R}^n : \alpha^\top x \geq \bar{\beta}_q^\alpha \left(\hat{S}^{r-1} \right) \right\},$$

where $\hat{S}^0 = S$ and $\bar{\beta}$ is defined as in (3.14).

Clearly, $\{\hat{S}^r\}_{r \geq 1}$ are polytopes with $m_0 + \sum_{i \in \mathcal{N}} m_i$ linear inequalities. Different from the infinite convergence of the quantile closure, we show that this approximate scheme converges finitely.

Theorem 12. *There exists a finite \hat{r} such that $\hat{S}^{\hat{r}} = \lim_{r \rightarrow \infty} \hat{S}^r$.*

Proof. Since $\{\hat{S}^r\}$ is a monotone non-increasing set sequence (i.e., $\hat{S}^0 \subseteq \hat{S}^1 \subseteq \dots \subseteq \hat{S}^r \subseteq \dots$), it is sufficient to show that there exists a $\hat{r} \geq 1$ such that $\hat{S}^{\hat{r}} = \hat{S}^{\hat{r}-1}$. Suppose that for any $r \geq 1$ it holds that $\hat{S}^r \subsetneq \hat{S}^{r-1}$, then from Proposition 17 there must exist $\hat{\alpha} \in \{D_{t\cdot}\}_{t \in [m_0]} \cup \{A_{t\cdot}^i\}_{t \in [m_i], i \in \mathcal{N}}$ such that $\bar{\beta}_q^{\hat{\alpha}}(\hat{S}^r) - \bar{\beta}_q^{\hat{\alpha}}(\hat{S}^{r-1}) \geq \frac{1}{(2^\phi - 1)2^\phi} \geq 2^{-2\phi}$. Since

$\left| \bar{\beta}_q^{\hat{\alpha}}(\hat{S}^r) \right| \leq 2^{\phi+\psi}$ by Proposition 17 and there are only $m_0 + \sum_{i \in \mathcal{N}} m_i$ choices of α , when $r > 2^{1+3\phi+\psi} (m_0 + \sum_{i \in \mathcal{N}} m_i)$, we must have $\hat{S}^r = \hat{S}^{r-1}$. \square \square

3.5.2 Approximation error estimation for covering CCPs

Considering a covering CCP, $\mathcal{X}^i = \{x \in S : (a^i)^\top x \geq 1\}$ and $p_i = 1/N$ for all $i \in [N]$, and $S = [0, M]^n$ with $M \geq \max_{i \in \mathcal{N}, j \in [n]: a_j^i \neq 0} \frac{1}{a_j^i}$. From [82] we know that the first quantile closure is equivalent to

$$S^1 = \bigcap_{\mathbf{g} \in \mathcal{G}} \left\{ x \in S : a_{\mathbf{g}}^\top x \geq 1 \right\},$$

where $(a_{\mathbf{g}})_j = \max_{i \in \mathbf{g}} a_j^i, \forall j \in [n]$, $\mathcal{G} = \{\mathbf{g} \subseteq [N] : |\mathbf{g}| = k+1\}$ and $k = \lfloor \epsilon N \rfloor$. Next we show a similar representation of the approximated quantile closure \hat{S}^1 .

Proposition 18. *For a covering CCP,*

$$\hat{S}^1 = \left\{ x \in S : (a^i)^\top x \geq \bar{\beta}_q^{a^i}(S), \forall i \in \mathcal{N} \right\}, \quad (3.15)$$

where $\bar{\beta}_q^{a^i}(S)$ is the $(k+1)$ th largest value among $\left\{ \min_{j \in [n]} \frac{a_j^i}{a_j^t} \right\}_{t \in \mathcal{N}}$.

Proof. Note that when $\alpha = a^i$, the optimal value of scenario $t \in \mathcal{N}$ is equal to $\beta_t^{a^i}(S) = \min_{j \in [n]} \frac{a_j^i}{a_j^t}$. Thus, the quantile bound is the $(k+1)$ th largest value among $\left\{ \min_{j \in [n]} \frac{a_j^i}{a_j^t} \right\}_{t \in \mathcal{N}}$ by the definition. \square \square

From the definition of approximation scheme, we have $S^1 \subseteq \hat{S}^1$. The following result measures the approximation error between S^1 and \hat{S}^1 .

Proposition 19. *For a covering CCP, $d_H(S^1, \hat{S}^1) \leq \left(\frac{\bar{a}}{\underline{a}} - 1 \right) \max_{i \in \mathcal{N}} \frac{\bar{\beta}_q^{a^i}}{\|a^i\|_2}$. where $\bar{a} = \min_{i \in \mathcal{N}} \max_{j \in [n]} a_j^i / \bar{\beta}_q^{a^i}(S)$ and $\underline{a} = \min_{\mathbf{g} \in \mathcal{G}} \min_{i \in \mathbf{g}} (a_{\mathbf{g}})_i$.*

Proof. (1) First of all, we would like to find a $\theta \geq 1$ such that

$$\hat{S}^1(\theta) \subseteq S^1 \subseteq \hat{S}^1$$

where

$$\hat{S}^1(\theta) = \left\{ x \in S : (a^i)^\top x \geq \theta \bar{\beta}_q^{a^i}(S), \forall i \in \mathcal{N} \right\}. \quad (3.16)$$

We make the following claim.

Claim: $\theta = \frac{\bar{a}}{\underline{a}}$ Satisfies (3.16), where $\bar{a} = \min_{i \in \mathcal{N}} \max_{j \in [n]} a_j^i / \bar{\beta}_q^{a^i}(S)$ and $\underline{a} = \min_{\mathbf{g} \in \mathcal{G}} \min_{i \in \mathbf{g}} (a_{\mathbf{g}})_i$. \square

Proof. First, we observe that $a_{\mathbf{g}} \geq \underline{a}\mathbf{e}$ for all $\mathbf{g} \in \mathcal{G}$, where \mathbf{e} is an all-one vector. Thus,

$$T := \{x \in S : \underline{a}\mathbf{e}^\top x \geq 1\} \subseteq S^1.$$

We all need to show is $\widehat{S}^1(\theta) \subseteq T$. From the definition of \bar{a} , there exists an $i_0 \in \mathcal{N}$ such that $\bar{a} = \max_{j \in [n]} a_j^{i_0} / \beta_q^{a^{i_0}}(S)$. Hence,

$$\widehat{S}^1(\theta) \subseteq \left\{x \in S : (a^{i_0})^\top x \geq \theta \beta_q^{a^{i_0}}(S)\right\} \subseteq \left\{x \in S : \bar{a}\mathbf{e}^\top x \geq \theta\right\} = T,$$

where the first inclusion is due to $i_0 \in \mathcal{N}$, the second inclusion comes from the definition of \bar{a} , and the third inequality is because of $\theta = \frac{\bar{a}}{\underline{a}}$. \diamond \square

(2) To prove our main result, we first note that

$$d_H(S^1, \widehat{S}^1) \leq d_H(\widehat{S}^1(\theta), \widehat{S}^1),$$

because of $\widehat{S}^1(\theta) \subseteq S^1$. Next, we show that

$$d_H(\widehat{S}^1(\theta), \widehat{S}^1) \leq (\theta - 1) \max_{i \in \mathcal{N}} \frac{\bar{\beta}_q^{a^i}}{\|a^i\|_2}.$$

This is because for any $x \in \widehat{S}^1 \setminus \widehat{S}^1(\theta)$, there exists $i_1 \in \mathcal{N}$ such that $\beta_q^{a^{i_1}} \leq (a^{i_1})^\top x \leq \theta \beta_q^{a^{i_1}}$ (otherwise $d(x, \widehat{S}^1(\theta)) = 0$), hence $d(x, \widehat{S}^1(\theta)) \leq (\theta - 1) \frac{\bar{\beta}_q^{a^{i_1}}}{\|a^{i_1}\|_2}$, which implies that

$$d_H(\widehat{S}^1(\theta), \widehat{S}^1) = \max_{x \in \widehat{S}^1 \setminus \widehat{S}^1(\theta)} d(x, \widehat{S}^1(\theta)) \leq (\theta - 1) \max_{i \in \mathcal{N}} \frac{\bar{\beta}_q^{a^i}}{\|a^i\|_2}.$$

\square

\square

Suppose that for each $i \in \mathcal{N}$, a^i is lower and upper bounded by $\underline{M}\mathbf{e}, \overline{M}\mathbf{e}$, respectively, i.e. $\overline{M}\mathbf{e} \geq a^i \geq \underline{M}\mathbf{e}$. Then we have $\beta_q^{a^i}(S) \geq \underline{M}/\overline{M}$ and $\underline{M}\sqrt{n} \leq \|a^i\|_2$ for each $i \in$

\mathcal{N} . It then follows that $\bar{a} \leq \bar{M}$ and $\underline{a} \geq \underline{M}/\bar{M}^2$. Therefore, by Proposition 19, the distance between first quantile closure S^1 and its approximation \hat{S}^1 is upper bounded as $d_H(S^1, \hat{S}^1) \leq \left(\frac{\bar{M}^2}{\underline{M}^2} - 1\right) \frac{\bar{M}^2}{\underline{M}^2 \sqrt{n}}$. Thus, in this setting, the approximation error vanishes asymptotically with problem dimension, i.e. $\lim_{n \rightarrow \infty} d_H(S^1, \hat{S}^1) = 0$.

We close this section by showing the approximation error for Example 3. For the given data we have $\bar{a} = 1, \underline{a} = 0.5$, $\left(\frac{\bar{a}}{\underline{a}} - 1\right) \max_{i \in \mathcal{N}} \frac{\bar{\beta}_a^i}{\|a^i\|_2} = \frac{\sqrt{2}}{4}$. The sets $\hat{S}^1, S^1, \hat{S}^1(\frac{\bar{a}}{\underline{a}})$ are shown in Figure 4.

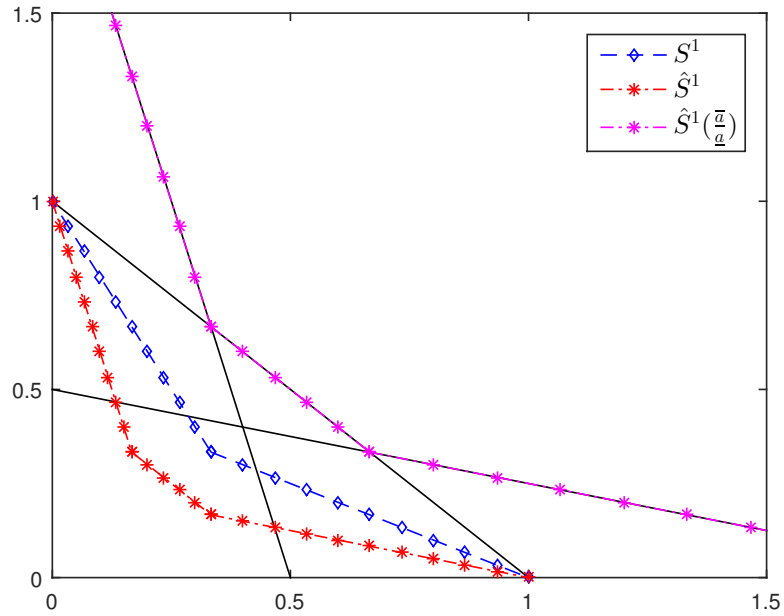


Figure 4: Illustration of the approximation error in Example 3

3.6 Generalized quantile closure

In this section, we generalize the quantile closure by grouping every $\kappa \in \mathbb{Z}_{++}$ scenarios together, where κ is no larger than cardinality of the smallest feasible set in \mathcal{Z} (i.e., $\kappa \leq \min_{C \in \mathcal{Z}} |C|$). Since we now simultaneously consider multiple scenarios while computing the quantile bound, this generalization can generate stronger quantile cuts. Note that for a given κ , there are $(N)_\kappa := \binom{N}{\kappa}$ different groups, denoted as $\binom{\mathcal{N}}{\kappa} := \{(\omega_i)_\kappa\}_{i \in [(N)_\kappa]}$, where $(\omega_i)_\kappa$ represents the i th group. We define the κ -quantile bound and cut as below.

Definition 5. Given $\alpha \in \mathbb{R}^n$ let $\{\beta_{i\kappa}^\alpha(S)\}_{i \in [(N)_\kappa]}$ be the optimal values of

$$\beta_{i\kappa}^\alpha(S) = \min \left\{ \alpha^\top x : x \in S, x \in \mathcal{X}^j, j \in (\omega_i)_\kappa \right\} \quad \forall i \in [(N)_\kappa]. \quad (3.17)$$

The κ -quantile $\beta_{q\kappa}^\alpha(S)$ is given by

$$\beta_{q\kappa}^\alpha(S) := \min_{\mathcal{C} \in \mathcal{Z}} \max_{i \in [(N)_\kappa]: (\omega_i)_\kappa \subseteq \mathcal{C}} \beta_{i\kappa}^\alpha(S) \quad (3.18)$$

and the associated “ κ -quantile cut” is

$$\alpha^\top x \geq \beta_{q\kappa}^\alpha(S). \quad (3.19)$$

Note that $\beta_{q\kappa}^\alpha(S)$ is a valid lower bound of $\min_{x \in X} \{\alpha^\top x\}$, since in (3.17), we simply choose the smallest bound among all the possible feasible subsets. In addition, when $p_i = \frac{1}{N}$ for all $i \in \mathcal{N}$, $\beta_{q\kappa}^\alpha(S)$ is the $((N)_\kappa - \binom{N-k}{\kappa} + 1)$ st largest value among $\{\beta_{i\kappa}^\alpha(S)\}_{i \in [(N)_\kappa]}$ with $k = \lfloor \epsilon N \rfloor$.

Next, we generalize Definition 2 to define κ -quantile closure.

Definition 6. The r th round κ -quantile closure S^r is defined as

$$S_\kappa^r := \bigcap_{\alpha \in \mathbb{R}^n} \left\{ x \in \mathbb{R}^n : \alpha^\top x \geq \beta_{q\kappa}^\alpha(S^{r-1}) \right\}, \quad r \geq 1.$$

We remark that when $\kappa = 1$, we recover Definition 1. The following results establish the validity of κ -quantile cut.

Proposition 20. For any α , $\alpha^\top x \geq \beta_{q\kappa}^\alpha(S)$ is valid for $\text{conv}(X)$.

Proof. Given $x \in X$, then there exists a feasible set $\widehat{\mathcal{C}} \in \mathcal{Z}$ such that x is contained in $S \cap_{i \in \widehat{\mathcal{C}}} \mathcal{X}^i$. We claim that there must exist a $\widehat{i} \in [(N)_\kappa]$, such that $(\omega_{\widehat{i}})_\kappa \subseteq \widehat{\mathcal{C}}$ and $\beta_{\widehat{i}\kappa}^\alpha(S) \geq \beta_{q\kappa}^\alpha(S)$; otherwise,

$$\max_{i \in [(N)_\kappa]: (\omega_i)_\kappa \subseteq \widehat{\mathcal{C}}} \beta_{i\kappa}^\alpha(S) < \beta_{q\kappa}^\alpha(S),$$

which implies existence of a strictly smaller κ -quantile bound, contradiction. Hence, we have

$$\alpha^\top x \geq \max_{i \in [(N)_\kappa]: (\omega_i)_\kappa \subseteq \widehat{\mathcal{C}}} \beta_{i\kappa}^\alpha(S) \geq \beta_{\widehat{i}\kappa}^\alpha(S) \geq \beta_{q\kappa}^\alpha(S).$$

□

□

Clearly, $\{S_\kappa^r\}_r$ is a nonincreasing set sequence with respect to r for any fixed κ . Next, we show that $\{S_\kappa^r\}_\kappa$ is a nonincreasing set sequence with respect to κ for any fixed r . First we observe that for any given α , $\{\beta_{q_\kappa}^\alpha(S)\}_\kappa$ is a nondecreasing sequence.

Lemma 5. *For any given α , $\beta_{q(\kappa-1)}^\alpha(S) \leq \beta_{q_\kappa}^\alpha(S)$ with $2 \leq \kappa \leq \min_{\mathcal{C} \in \mathcal{Z}} |\mathcal{C}|$.*

Proof. Let \mathcal{C}_κ be one of the feasible sets that achieve κ -quantile defined in (3.18). Define

$$(\omega_{\hat{i}})_{\kappa-1} := \arg \max_{i \in [N_{\kappa-1}]: (\omega_i)_{\kappa-1} \subseteq \mathcal{C}_\kappa} \beta_{i(\kappa-1)}^\alpha(S).$$

Since $\kappa \leq \min_{\mathcal{C} \in \mathcal{Z}} |\mathcal{C}| \leq |\mathcal{C}_\kappa|$, there exists at least one $(\omega_{\hat{i}})_{\kappa-1} \subseteq \mathcal{C}_\kappa$ with cardinality κ , which contains set $(\omega_{\hat{i}})_{\kappa-1}$. Thus,

$$\begin{aligned} \beta_{q_\kappa}^\alpha(S) &= \max_{i \in [(N)_{\kappa}]: (\omega_i)_\kappa \subseteq \mathcal{C}_\kappa} \beta_{i\kappa}^\alpha(S) \geq \beta_{i\kappa}^\alpha(S) \geq \beta_{i(\kappa-1)}^\alpha(S) \\ &= \max_{i \in [N_{\kappa-1}]: (\omega_i)_{\kappa-1} \subseteq \mathcal{C}_\kappa} \beta_{i(\kappa-1)}^\alpha(S) \geq \min_{\mathcal{C} \in \mathcal{Z}} \max_{i \in [N_{\kappa-1}]: (\omega_i)_{\kappa-1} \subseteq \mathcal{C}} \beta_{i(\kappa-1)}^\alpha(S) := \beta_{q(\kappa-1)}^\alpha(S). \end{aligned}$$

where the first equality is the definition of \mathcal{C}_κ , the first inequality comes from $(\omega_{\hat{i}})_{\kappa-1} \subseteq \mathcal{C}_\kappa$ with cardinality κ , the second inequality is due to $(\omega_{\hat{i}})_{\kappa-1} \subset (\omega_i)_{\kappa-1}$, the second equality is the definition of $(\omega_{\hat{i}})_{\kappa-1}$, the third inequality is due to $\mathcal{C}_\kappa \in \mathcal{Z}$, while the last equality is the definition of $\kappa - 1$ quantile $\beta_{q(\kappa-1)}^\alpha(S)$. \square \square

Proposition 21. *For any fixed $r \in \mathbb{Z}_{++}$, the set sequence $\{S_\kappa^r\}_\kappa$ is nonincreasing.*

Proof. It is sufficient to show that $S_\kappa^r \subseteq S_{\kappa-1}^r$ for any $2 \leq \kappa \leq \min_{\mathcal{C} \in \mathcal{Z}} |\mathcal{C}|$. We prove it by induction on r . When $r = 1$, by Lemma 5, the quantile cut $\alpha^\top x \geq \beta_{q(\kappa-1)}^\alpha(S)$ is dominated by $\alpha^\top x \geq \beta_{q_\kappa}^\alpha(S)$ for any given α . Thus, $S_\kappa^1 \subseteq S_{\kappa-1}^1$. Suppose that for any $\gamma \leq r$, we have $S_\kappa^\gamma \subseteq S_{\kappa-1}^\gamma$. Now let $\gamma = r + 1$. Note that

$$\beta_{q(\kappa-1)}^\alpha(S_\kappa^r) \geq \beta_{q(\kappa-1)}^\alpha(S_{\kappa-1}^r)$$

due to the hypothesis. Replacing S by set S_κ^r in Lemma 5, we have

$$\beta_{q_\kappa}^\alpha(S_\kappa^r) \geq \beta_{q(\kappa-1)}^\alpha(S_\kappa^r) \geq \beta_{q(\kappa-1)}^\alpha(S_{\kappa-1}^r).$$

Hence, the quantile cut $\alpha^\top x \geq \beta_{q(\kappa-1)}^\alpha(S_{\kappa-1}^r)$ is dominated by $\alpha^\top x \geq \beta_{q_\kappa}^\alpha(S_\kappa^r)$ for any given α . This implies that $S_\kappa^{r+1} \subseteq S_{\kappa-1}^{r+1}$. \square \square

The next definition is a generalization of Definition 3.

Definition 7. A set $\mathbf{g}_\kappa \subseteq \binom{\mathcal{N}}{\kappa}$ is a “ κ -partial covering subset” if it “intersects” with all of feasible scenario subsets in \mathcal{Z} , i.e., for any $\widehat{\mathcal{C}} \in \mathcal{Z}$, there exists a $\omega \in \mathbf{g}_\kappa$ we have $\omega \subseteq \widehat{\mathcal{C}}$. Also, a set \mathbf{g}_κ is a “minimal” κ -partial covering subset if there does not exist another κ -partial covering subset $\mathbf{g}'_\kappa \subseteq \mathcal{N}$ such that $\mathbf{g}'_\kappa \subsetneq \mathbf{g}_\kappa$. We let \mathcal{G}_κ denote the collection of all of the minimal partial covering subsets.

Example 4. Suppose $\mathcal{N} = \{1, 2, 3, 4\}$ with $p_i = \frac{1}{4}$ for all $i \in \mathcal{N}$ and $\epsilon = 0.25$ ($k := \lfloor \epsilon N \rfloor = 1$). In this case, $\mathcal{Z} = \{\{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}\}$. Let $\kappa = 2$ and $\mathbf{g}_2 = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ be a 2-partial covering subset since for each feasible scenario subset $\mathcal{C} \in \mathcal{Z}$, there exists an element $\omega \in \mathbf{g}_\kappa$, which is also contained in \mathcal{C} . Note that \mathbf{g}_2 is also minimal since there is no other partial covering subset which has smaller size and is a subset of \mathbf{g}_2 .

Similar to Theorems 9 and 10, the following two theorems demonstrate the characterization of r th κ quantile closure using \mathcal{G}_κ defined above, and the convergence of κ quantile closure sequence. The proofs are nearly identical to those for Theorems 9 and 10 and are omitted here.

Theorem 13. For any $r \in \mathbb{Z}_{++}$ and $2 \leq \kappa \leq \min_{\mathcal{C} \in \mathcal{Z}} |\mathcal{C}|$

$$S_\kappa^r = \bigcap_{\mathbf{g}_\kappa \in \mathcal{G}_\kappa} \text{conv} \left[\bigcup_{\omega \in \mathbf{g}_\kappa} \left(S_\kappa^{r-1} \bigcap_{i \in \omega} \mathcal{X}^i \right) \right],$$

where $S_\kappa^0 = S$.

Theorem 14. For each $2 \leq \kappa \leq \min_{\mathcal{C} \in \mathcal{Z}} |\mathcal{C}|$, the set sequence $\{S_\kappa^r\}$ converges to $\text{conv}(X)$ with respect to the Hausdorff distance; i.e., $\bar{S}_\kappa = \lim_{r \rightarrow \infty} S_\kappa^r = \text{conv}(X)$.

We close this section by remarking about the exactness of κ -quantile bound. Suppose each scenario occurs with equal probability, i.e. $p_i = \frac{1}{N}$ for each $i \in \mathcal{N}$, and let $k = \lfloor \epsilon N \rfloor$ where $\epsilon = \frac{r}{N}$ for a constant r that is independent of N . Let $\kappa = N - k$, and the quantile bound $\beta_{q(N-k)}^c(S)$ is the optimal value to (3.1), since $(N)_\kappa$ equals to the collection of all the feasible subsets and in this case, the quantile bound is equal to the minimum value

among all the feasible subsets. Since $k = O(1)$, to compute the quantile bound $\beta_{q(N-k)}^c(S)$, we only need to solve a polynomial number $((\binom{N}{N-k}) = (O(N^k)))$ of optimization problems (3.17) with $\alpha = c$. Note that this fact does not hold for general k , for example, if $\epsilon = \frac{1}{2}$ (i.e., $k = \frac{1}{2}N$), clearly, to evaluate the quantile bound $\beta_{q(N-k)}^c(S)$, one has to solve $\Omega(2^N/\sqrt{N})$ number of optimization problems.

3.7 Separation over the first quantile closure

3.7.1 Complexity

We first show that separating over the first quantile closure even in the absence of integrality restrictions is NP-hard. Our proof is based on the constructions in [68] and [82].

Theorem 15. *The separation over S^1 is, in general, NP-hard.*

Proof. We consider a covering CCP where $\mathcal{X}^i = \{x \in S : (a^i)^\top x \geq 1\}$ and $p_i = 1/N$ for all $i \in [N]$, and $S = [0, M]^n$ with $M \geq \max_{i \in \mathcal{N}, j \in [n]: a_j^i \neq 0} \frac{1}{a_j^i}$. From [82] it can be shown that

$$S^1 = \bigcap_{\mathbf{g} \in \mathcal{G}} \left\{ x \in S : a_{\mathbf{g}}^\top x \geq 1 \right\}, \quad (3.20)$$

where $(a_{\mathbf{g}})_j = \max_{i \in \mathbf{g}} a_j^i, \forall j \in [n]$, $\mathcal{G} = \{\mathbf{g} \subseteq [N] : |\mathbf{g}| = k + 1\}$ and $k = \lfloor \epsilon N \rfloor$ (see Definition 3).

For a given solution $\hat{x} \in S$, to separate it from S^1 is equivalent to solving the following problem

$$\delta^* = \min_{\mathbf{g} \in \mathcal{G}} \max_{i \in \mathbf{g}} \sum_{j \in [n]} a_j^i \hat{x}_j - 1, \quad (3.21)$$

i.e., find a violated constraint of the form $a_{\mathbf{g}}^\top x \geq 1$ in the description (3.20). If $\delta^* < 0$, then $\hat{x} \notin S^1$; otherwise, $\hat{x} \in S^1$. Consider the decision version of this separation problem:

(SepCCP) Given nonnegative integers $\{a_j^i\}_{i \in \mathcal{N}, j \in [n]}$ and a rational vector $\hat{x} \in S$, does there exist a $\mathbf{g} \subseteq \mathcal{N}$ with $|\mathbf{g}| = k + 1 (k < N)$ such that $\sum_{j \in [n]} \max_{i \in \mathbf{g}} a_j^i \hat{x}_j < 1$?

Following [68] we can show that SepCCP is NP-complete via reduction from the NP-complete problem CLIQUE which asks

(CLIQUE) Given a graph with nodes V and edges E , does it contain a clique of size C ?

Given an instance of CLIQUE we can construct an instance of SepCCP as $[n] = V, \mathcal{N} = E, \hat{x}_j = \frac{1}{C+1}$ for all $j \in [n]$, $k+1 = \frac{1}{2}C(C-1)$ and $a_j^i = 1$ if edge i contains nodes j and $a_j^i = 0$ otherwise. It is easy to verify that if CLIQUE has an answer Yes, then there exists a subgraph with edges $\mathbf{g} \subseteq \mathcal{N}$ and $|\mathbf{g}| = \frac{1}{2}C(C-1)$ such that

$$\sum_{j \in [n]} \max_{i \in \mathbf{g}} a_j^i \hat{x}_j = \frac{C}{C+1} < 1.$$

Hence, SepCCP has an answer Yes. Conversely, if SepCCP has an answer Yes, this implies that

$$\sum_{j \in [n]} \max_{i \in \mathbf{g}} a_j^i < C+1;$$

i.e., there exists a subgraph with edges $\mathbf{g} \subseteq \mathcal{N}$ and $|\mathbf{g}| = \frac{1}{2}C(C-1)$, which contains at most C nodes. Clearly, thus CLIQUE has an answer Yes. \square \square

Different from S^1 , separation over the first approximated quantile closure \hat{S}^1 is easy since for a given solution \hat{x} , we can verify whether $\hat{x} \in \hat{S}^1$ or not by simply comparing $\alpha^\top \hat{x}$ with $\bar{\beta}_q^\alpha(S)$ for all $\alpha \in \{D_t\}_{t \in [m_0]} \cup \{A_t^i\}_{t \in [m_i], i \in \mathcal{N}}$. If there exists a $\hat{\alpha} \in \{D_t\}_{t \in [m_0]} \cup \{A_t^i\}_{t \in [m_i], i \in \mathcal{N}}$ such that $\hat{\alpha}^\top \hat{x} < \bar{\beta}_q^{\hat{\alpha}}(S)$, then $\hat{x} \notin \hat{S}^1$ with a separating hyperplane $\hat{\alpha}^\top x \geq \bar{\beta}_q^{\hat{\alpha}}(S)$; otherwise, $\hat{x} \in \hat{S}^1$.

3.7.2 A heuristic separation algorithm

In this section, we introduce a heuristic separation algorithm over the first quantile closure S^1 inspired by Theorem 9. First of all, we relax the integrality of x variables, i.e., assume that $S \cap \mathcal{X}^i := \{x \in \mathbb{R}^n : \bar{G}_i(x) \leq 0\}$. Then we define a continuous relaxation set as $X^{con} := \{x \in \mathbb{R}^n : \bar{G}_i(x) \leq \bar{M}_i(1 - z_i), \sum_{i \in [N]} p_i z_i \geq 1 - \epsilon, z \in [0, 1]^N\}$ with appropriate $\{\bar{M}_i\}$, such that $\text{conv}(X) \subseteq X^{con}$.

Given an optimal solution \hat{x} which optimizes $c^\top x$ over set X^{con} , it is unlikely to be feasible to X . The proposed heuristic algorithm is to find a minimal partial cover over the subset of scenarios that does not contain \hat{x} . To select such a partial cover, we prioritize the

scenarios by norm of the constraint violations, i.e. we sort $\{\|\bar{G}_i(\hat{x})\|_\infty\}_{i \in \mathcal{N}}$ in a descending order such that $\|\bar{G}_{\sigma(1)}(\hat{x})\|_\infty \geq \dots \geq \|\bar{G}_{\sigma(N)}(\hat{x})\|_\infty$ where σ is a permutation of \mathcal{N} , then select the scenarios according to this order until total probability mass is strictly greater than ϵ . Let $\nu := \min\{j \in \mathcal{N} : \sum_{i=1}^j p_{\sigma(i)} > \epsilon\}$ and $\mathbf{g} := \{\sigma(i)\}_{i \in [\nu]}$. By Theorem 9, if \hat{x} can be separated from $\text{conv}\left[\bigcup_{i \in \mathbf{g}} (S \cap \mathcal{X}^i)\right]$, then it can be also separated from S^1 . To separate \hat{x} , we consider the following optimization problem that minimizes 2-norm distance between set $\text{conv}\left[\bigcup_{i \in \mathbf{g}} (S \cap \mathcal{X}^i)\right]$ and point \hat{x} :

$$\delta^* = \min_{x, \lambda} \|\hat{x} - x\|_2^2 \quad (3.22a)$$

$$\text{s.t. } \lambda_i \bar{G}_i(x^i / \lambda_i) \leq 0, \quad i \in \mathbf{g}, \quad (3.22b)$$

$$\sum_{i \in \mathbf{g}} x^i = x, \quad (3.22c)$$

$$\sum_{i \in \mathbf{g}} \lambda_i = 1, \quad (3.22d)$$

$$\lambda_i \geq 0, \forall i \in \mathbf{g}. \quad (3.22e)$$

where $\lambda_i \bar{G}_i(x^i / \lambda_i)$ is the perspective function associated with \bar{G} and is jointly convex in (λ_i, x^i) . The system (3.22b) - (3.22e) provides an equivalent reformulation of $\text{conv}[\bigcup_{i \in \mathbf{g}} (S \cap \mathcal{X}^i)]$ (see [27] for details). Note that for each $i \in \mathcal{N}$, if $\bar{G}_i(x)$ is second order cone representable, then (3.22b) is second order cone. The separation problem (3.22) is a convex optimization problem, which is relatively easy to solve. Given an optimal solution (x^*, λ^*) of (3.22) with optimal objective value δ^* , we check if $\delta^* > 0$, then $\hat{x} \notin S^1$ with a separating hyperplane $(\hat{x} - x^*)^\top (x - x^*) \leq 0$, which is valid for all $x \in S^1$. Thus, we can add this valid inequality to X^{con} and repeat the above steps until we cannot separate any more. The detailed procedure is described in Algorithm 3. The output (i.e., set E) can be directly added to (3.4).

We remark that Algorithm 3 can be applied to the generalized quantile closure. However, it might be expensive to enumerate all the κ -groups in $\binom{[N]}{\kappa}$ when $\kappa \geq 2$. Suppose that κ is a divisor of N , then we can consider the partition \bar{G}_κ of set $[N]$ by evenly dividing set $[N]$ into $\frac{N}{\kappa}$ groups, i.e., $\bar{G}_\kappa = \{\mathbf{g}_j\}_{j \in [N/\kappa]}$ with $\mathbf{g}_j = \{(j-1)\kappa + 1, \dots, j\kappa\}$. Next, we let $\hat{p}_j = \min_{i \in \mathbf{g}_j} p_i$ for each $j \in [N/\kappa]$. Then by Proposition 1 in [4], we have that the following

Algorithm 3 A heuristic separation algorithm.

```

1: Let  $E = \emptyset$ ,  $\delta^* = \infty$  and  $\underline{\delta} > 0$  be a tolerance parameter.
2: while true do
3:   Let  $\hat{x} \in \arg \min_{x \in X^{con} \cap E} c^\top x$  be an optimal solution.
4:   Sort  $\{\|\bar{G}_i(\hat{x})\|_\infty\}_{i \in \mathcal{N}}$  in a descending order such that  $\|\bar{G}_{\sigma(1)}(\hat{x})\|_\infty \geq \dots \geq \|\bar{G}_{\sigma(N)}(\hat{x})\|_\infty$ 
5:   Let  $\nu := \min\{j \in \mathcal{N} : \sum_{i=1}^j p_{\sigma(i)} > \epsilon\}$  and  $\mathbf{g} := \{\sigma(i)\}_{i \in [\nu]}$ 
6:   if  $\nu$  does not exist then
7:     return Set  $E$ .
8:   else
9:     Solve (3.22) with an optimal solution  $(x^*, \lambda^*)$  and optimal objective value  $\delta^*$ 
10:    if  $\delta^* > \underline{\delta}$  then
11:      Add  $(\hat{x} - x^*)^\top (x - x^*) \leq 0$  to set  $E$ 
12:    else
13:      return Set  $E$ .
14:    end if
15:  end if
16: end while

```

CCP yields a relaxation of (3.4)

$$\min_{x \in S, \hat{z} \in \{0,1\}^{N/\kappa}} \left\{ c^\top x : x \in S, G_i(x) \leq M_i(1 - \hat{z}_j), \forall i \in \mathbf{g}_j, \forall j \in [N/\kappa], \sum_{j \in [N/\kappa]} \hat{p}_j \hat{z}_j \geq 1 - \hat{\epsilon} \right\}, \quad (3.23)$$

where $\hat{\epsilon} := 1 + \epsilon - \sum_{j \in [N/\kappa]} \hat{p}_j$. Now we can apply Algorithm 3 to the relaxed CCP (3.23).

3.8 Numerical illustration

In this section, we present a numerical study to illustrate Algorithm 3 and the strength of quantile cuts. We consider the following norm optimization problem, which has also been studied in [48, 103],

$$\min_{x, z} \left\{ c^\top x : x \in [0, 100]^n, \sum_{i \in \mathcal{N}} p_i \mathbb{I} \left[\sum_{j=1}^n \xi_j^i x_j^2 \leq 100 \right] \geq 1 - \epsilon \right\}, \quad (3.24)$$

where the support $\{\xi^i\}_{i=1}^n$ of $\boldsymbol{\xi}$ are non-negative. We consider the cases when $n \in \{10, 20, 30\}$, $N \in \{60, 80, 100\}$, $\epsilon \in \{0.05, 0.10, 0.15\}$, and each scenario occurs with equal probability, i.e. $p_i = \frac{1}{N}$. The cost vector c and unknown data $\{\xi^i\}_{i \in \mathcal{N}}$ are randomly generated, where each component of the cost vector c is integral and randomly distributed between -10 and -1 with equal probability, and ξ_j^i is also integral and uniformly distributed

between 1 and 99. Under this setting, the MINLP reformulation (3.4) is

$$\min_{x,z} \left\{ c^\top x : x \in [0, 100]^n, \sum_{j=1}^n \xi_j^i x_j^2 \leq 100 + (M_i - 100)(1 - z_i), \forall i \in \mathcal{N}, z \in Z \right\}, \quad (3.25)$$

where $M_i = 100 \sum_{j=1}^n \xi_j^i$ suffices for each $i \in \mathcal{N}$. In this case, the separation problem (3.22) is equivalent to a second order cone program

$$\delta^* = \min_{x,\lambda} \|\hat{x} - x\|_2^2 \quad (3.26a)$$

$$\text{s.t.} \quad \sum_{j=1}^n \xi_j^i (x_j^i)^2 \leq 100\lambda_i^2, \quad i \in \mathbf{g}, \quad (3.26b)$$

$$\sum_{i \in \mathbf{g}} x^i = x, \quad (3.26c)$$

$$\sum_{i \in \mathbf{g}} \lambda_i = 1, \quad (3.26d)$$

$$\lambda_i \geq 0, \forall i \in \mathbf{g}. \quad (3.26e)$$

We compare three solution approaches. The first one is to run Algorithm 3 and add all separated quantile cuts to (3.25), the second is to run Algorithm 3 to separate generalized quantile cuts for $\kappa = 2$ for the relaxation (3.23), and finally to use the commercial solver CPLEX to solve the formulation without any quantile cuts. The overall time limit is set to be 4 hours. The results are listed in Table 7, where the optimality gap is computed as

$$|\text{best upper bound}|/|\text{best lower bound}| - 1.$$

We use $t_{\text{sep}}, t_{\text{C}}, t_{\text{total}}$ to denote the running time of Algorithm 3, the running time of solver CPLEX and overall running time, respectively, and use g_r and $g_e(\%)$ to denote the optimality gap at root node and the best optimality gap when the solution procedure is completed. All instances were executed on a laptop with a 2.67 GHz processor and 4GB RAM, while CPLEX 12.5.1 was used with its default setting.

In Table 7, we observe that the quantile cuts separated using Algorithm 3 can reduce the root gap from more than 450% to within 45%, while the 2-quantile cuts further reduce the root gap to within 25%. CPLEX with quantile cuts can solve 22 out of 27 instances within time limit, and with 2-quantile cuts can solve 23 out of 27, while CPLEX without

Table 7: Performance of CPLEX with and without quantile cuts

N	ϵ	n	With quantile cuts					With $\kappa = 2$ - quantile cuts					Without quantile cuts		
			t_{sep}	t_C	t_{total}	$g_r(\%)$	$g_e(\%)$	t_{sep}	t_C	t_{total}	$g_r(\%)$	$g_e(\%)$	t_{total}	$g_r(\%)$	$g_e(\%)$
60	0.05	10	10.8	6.3	17.2	8.3	0.0	41.8	5.3	47.1	5.5	0.0	42.9	463.7	0.0
60	0.05	20	158.7	44.0	202.7	13.0	0.0	948.9	33.0	981.9	8.4	0.0	153.5	681.7	0.0
60	0.05	30	1620.4	374.2	1994.6	11.1	0.0	4123.2	385.6	4508.8	9.3	0.0	1544.4	890.9	0.0
60	0.1	10	8.4	7.4	15.8	20.2	0.0	54.6	8.9	63.5	5.5	0.0	31.5	669.8	0.0
60	0.1	20	159.1	109.1	268.1	19.2	0.0	854.2	69.9	924.2	10.1	0.0	653.4	962.6	0.0
60	0.1	30	628.1	4561.3	5189.4	27.9	0.0	1467.4	1773.6	3240.9	15.4	0.0	14402.6	1201.3	4.6
60	0.15	10	5.5	21.4	26.9	21.9	0.0	37.5	12.2	49.6	11.5	0.0	70.2	849.4	0.0
60	0.15	20	41.7	1770.3	1812.0	35.5	0.0	479.1	686.6	1165.6	15.5	0.0	1931.7	1208.3	0.0
60	0.15	30	777.0	8953.2	9730.2	28.1	0.0	1624.4	7467.9	9092.3	17.3	0.0	14400.7	1497.0	6.5
80	0.05	10	28.9	8.2	37.1	11.4	0.0	97.9	4.8	102.7	6.5	0.0	27.2	480.7	0.0
80	0.05	20	613.7	100.4	714.1	17.1	0.0	2652.4	85.7	2738.1	8.1	0.0	311.0	691.7	0.0
80	0.05	30	2231.1	558.5	2789.6	21.7	0.0	4454.8	779.2	5234.1	11.7	0.0	5999.9	852.1	0.0
80	0.1	10	15.4	12.9	28.3	10.0	0.0	89.2	16.5	105.7	7.4	0.0	91.8	704.8	0.0
80	0.1	20	342.8	497.1	839.9	22.0	0.0	1964.7	292.8	2257.5	11.3	0.0	3763.8	956.5	0.0
80	0.1	30	1488.7	12911.4	14400.0	29.8	1.5	3745.8	10654.5	14400.0	15.1	0.8	14400.0	1228.6	20.2
80	0.15	10	14.8	293.6	308.3	41.5	0.0	136.0	43.4	179.4	12.3	0.0	572.3	821.5	0.0
80	0.15	20	297.3	3286.6	3583.8	27.5	0.0	1714.9	1918.6	3633.5	14.6	0.0	14400.0	1200.2	4.5
80	0.15	30	1749.7	12650.8	14400.0	42.7	7.2	3999.9	10400.6	14400.0	17.2	4.4	14400.0	1493.7	29.4
100	0.05	10	27.1	7.7	34.9	13.2	0.0	170.9	8.6	179.5	7.8	0.0	27.8	473.1	0.0
100	0.05	20	647.7	124.4	772.2	19.2	0.0	1447.3	166.9	1614.2	11.9	0.0	433.3	681.4	0.0
100	0.05	30	3140.9	1766.3	4907.2	26.2	0.0	5432.8	956.1	6388.9	14.9	0.0	14400.0	843.2	4.1
100	0.1	10	19.3	29.9	49.2	15.0	0.0	178.3	21.3	199.6	9.5	0.0	136.6	684.0	0.0
100	0.1	20	690.2	5805.7	6495.9	27.1	0.0	2195.0	2178.2	4373.2	13.4	0.0	14400.0	998.2	0.8
100	0.1	30	3831.8	10568.4	14400.0	35.2	10.0	6362.9	8037.4	14400.0	15.8	7.3	14400.0	1219.5	16.8
100	0.15	10	29.4	100.5	129.9	18.1	0.0	117.3	97.1	214.4	14.3	0.0	1693.8	848.5	0.0
100	0.15	20	339.3	14062.2	14401.5	41.7	5.6	2723.0	9710.6	12433.7	15.9	0.0	14401.6	1224.8	12.1
100	0.15	30	3033.6	11366.8	14400.0	40.7	19.8	3736.6	10663.8	14400.0	23.0	14.3	14401.9	1483.4	42.4

the cuts can only solve 17 out of 27. For the unsolved instances, the average remaining gaps (8.8%, 6.8%) for the first two methods are also much smaller than for the third (14.2%). Comparing the first and second methods with the third one, we see that the quantile cuts help reduce nearly more than half of the solution time for those instances that can be solved within time limit. Comparing the total running time for first two methods with that of the third one, there are still 22 out of 27 and 14 out of 27 instances which take a shorter time or smaller ending gap for the first method. These results demonstrate the effectiveness of quantile cuts and 2- quantile cuts separated using Algorithm 3 for solving a convex chance constrained problem. We also note that the separation time of first method with quantile cuts is usually shorter than the second one with 2-quantile cuts. However, the second method has a much smaller root gap, therefore, it can solve more instances or have smaller ending gaps. Hence, we recommend to use κ -quantile cuts over regular quantile cuts to solve the large-scale instances.

CHAPTER IV

ON BICRITERIA APPROXIMATION OF CHANCE-CONSTRAINED COVERING PROBLEMS

4.1 *Introduction*

4.1.1 Problem setting

A chance constrained stochastic program (CCSP) involves optimization over constraints (specified by stochastic data) which are required to be satisfied with a prescribed probability level. A CCSP can be formulated as

$$v_\epsilon^* = \min_x \left\{ c^\top x : x \in S, \mathbb{P}\{\boldsymbol{\xi} : x \notin X(\boldsymbol{\xi})\} \leq \epsilon \right\}. \quad (4.1)$$

In the above formulation, $\boldsymbol{\xi}$ denotes a random data vector, and $X(\boldsymbol{\xi})$ denotes a system of stochastic constraints whose data is specified by the random vector $\boldsymbol{\xi}$. The set $S \subseteq \mathbb{R}^n$ is a system of deterministic constraints. The CCSP (4.1) seeks a solution $x \in S$ that minimize the cost $c^\top x$ and violates the stochastic constraints $X(\boldsymbol{\xi})$ with probability at most ϵ where $\epsilon \in (0, 1)$ is a prespecified risk level.

In this chapter, we consider a CCSP with uncertain covering constraints under finite support. In particular, we assume that

- a. $\boldsymbol{\xi}$ is a random vector with a finite distribution supported on $X\boldsymbol{i} = \{\xi^1, \dots, \xi^N\}$, where each ξ^i for $i \in \mathcal{N} := \{1, \dots, N\}$ corresponds to a scenario with a probability mass $\frac{1}{N}$;
- b. for each $i \in [N]$, the stochastic constraint system is of the form $X(\xi^i) = \{x \in S : A(\xi^i)x \geq b(\xi^i)\}$ with $A(\xi^i) \in \mathbb{R}_{++}^{m_i \times n}$, $b(\xi^i) \in \mathbb{R}_+^{m_i}$;
- c. the deterministic constraint system $S \subseteq C$ is nonempty, where C is a closed convex cone and for each $x \in S$ and positive scalar $\eta \geq 1$, $\eta x \in S$;
- d. the objective cost vector $c \in C^*$, where C^* is the dual cone of C ; and
- e. ϵN is an integer number.

Note that in Assumption a., equal probability mass is not necessary and a generalization will be shown in the subsequent sections. In Assumption b., nonnegativity of matrices $\{A(\xi^i)\}_{i \in [N]}$ and vectors $\{b(\xi^i)\}_{i \in [N]}$ defines the linear covering inequalities. Assumptions c. and d. are to guarantee that any feasible solution to (4.1) scaled by a number no smaller than 1 is still feasible and v_ϵ^* is bounded. Assumption e. is without loss of generality since we can always replace ϵN by $\lfloor \epsilon N \rfloor$. Under the above assumptions, problem (4.1) is feasible and bounded, and has a nonnegative optimal solution. For notational simplicity, we denote let $X^i = X(\xi^i)$, $b^i := b(\xi^i)$ and $A^i := A(\xi^i)$ for all $i \in [N]$. Thus, the stochastic constraint system for scenario i is $X^i = \{x \in S : A^i x \geq b^i\}$.

By introducing binary variables z_i for $i \in [N]$, then we can reformulate (4.1) as a mixed integer program (MIP) (c.f. [82, 120]):

$$v_\epsilon^* = \min_{x, z} \left\{ c^\top x : x \in S, A^i x \geq b^i z_i, \forall i \in [N], z \in Z_\epsilon \cap \mathbb{B}^N \right\}, \quad (4.2)$$

where

$$Z_\epsilon := \left\{ z \in [0, 1]^N : \sum_{i \in [N]} z_i \geq N - \epsilon N \right\}, \quad \mathbb{B} = \{0, 1\}.$$

Let X_ϵ be the feasible region of (4.2), i.e.

$$X_\epsilon := \{x : x \in S, A^i x \geq b^i z_i, \forall i \in [N], z \in Z_\epsilon \cap \mathbb{B}^N\}. \quad (4.3)$$

and \widehat{X}_ϵ be its continuous relaxation set, i.e.,

$$\widehat{X}_\epsilon := \{x : x \in S, A^i x \geq b^i z_i, \forall i \in [N], z \in Z_\epsilon\}. \quad (4.4)$$

4.1.2 Contributions

The covering CCSP problem described above has been shown to NP-hard [82]. In this chapter, we study approximation of CCSP with provable guarantees. In particular, we consider bicriteria approximations. Given a violation ratio $\sigma \geq 1$ and an optimality ratio $\gamma \geq 1$, a (σ, γ) -bicriteria approximation algorithm for CCSP (4.1) returns a solution $\widehat{x} \in S$, such that $\mathbb{P}\{\xi : \widehat{x} \notin X(\xi)\} \leq \sigma\epsilon$ and $c^\top \widehat{x} \leq \gamma v_\epsilon^*$, i.e., the solution violates the uncertain constraints with probability at most $\sigma\epsilon$ and has an objective value at most γ times the optimal value. As a special case, when the violation ratio $\sigma = 1$, we have a single criterion approximation

algorithm with an approximation ratio $\gamma \geq 1$. Note that σ, γ may be dependent on the risk parameter ϵ and underlying probability distribution \mathbb{P} .

In this chapter we make the following contributions regarding approximation of chance constraint covering problems with finite scenarios.

1. We show that some well known approximation approaches for covering CCSP can have arbitrarily bad approximation ratios when no constraint violations are allowed.
2. We prove that, unless $P=NP$, it is impossible to obtain a polynomial time algorithm with a constant factor approximation if no violations are allowed, i.e. if $\sigma = 1$. This motivates the need for bicriteria approximations.
3. We analyze a simple scaling approach and show that, given a violation ratio of $\sigma > 1$, the approach provides a solution within a factor of $\gamma = \sigma/(\sigma - 1)$ of the optimal value.
4. We prove that the analysis above is tight.
5. The proposed scaling approach scales a solution to the LP relaxation of the MIP formulation (4.2). We show that a tighter LP relaxation may not lead to a better approximation guarantee.

4.1.3 Related Literature

A variety of approximation approaches have been developed for general CCSP as well as for special cases. The papers [23], [73] and related works develop convex restrictions of the CCSP feasible region, optimizing over which provides a feasible solution. In [4], the authors developed a specialized heuristic for covering CCSP based on a convex relaxation. As shown in Section 4.2, these approaches do not come with any provable approximation guarantees. To the best of our knowledge, all existing approximation algorithms with provable guarantees have been proposed for chance constrained *combinatorial* optimization problem. For example, [41] proposed constant factor approximation algorithms for certain classes of chance constrained combinatorial covering problems. Subsequently, [42] developed a fully polynomial time approximation scheme for a chance constrained knapsack problem where the item sizes are drawn from independent normal distributions. In [104], the authors

studied a two-stage chance constrained set covering problem with budget constraint and proposed a polynomial time approximation algorithm by rounding linear program relaxation solutions. In contrast to these works, we consider approximations of chance constrained covering problems with continuous variables.

The remaining of the chapter is organized as follows. Section 4.2 shows the inapproximability results and three popular approximation algorithms which yield worse approximation results. Section 4.3 provides approximation results and shows the numerical study and Section 4.4 extends the results to a CCSP with general probability mass.

4.2 Inapproximability

We first show that three existing approximation approaches can have arbitrarily bad approximation ratios for covering CCSPs, and then provide a formal single-criterion inapproximability result.

4.2.1 CVaR approximation

A well known approximation of CCSP is to replace the nonconvex probabilistic constraint by a convex constraint defined by the conditional value at risk or **CVaR** (see [73] for details). For the covering CCSP considered in this chapter, the resulting formulation is

$$v_\epsilon^{\text{CVaR}} = \min_x \left\{ c^\top x : x \in S, \inf_\beta [-\epsilon\beta + \mathbb{E}(-a(x, \xi) + \beta)_+] \leq 0 \right\}, \quad (4.5)$$

where $(t)_+ = \max(t, 0)$ and $a(x, \xi) = \min_{j \in [m]} [A_{j\cdot}(\xi)x - b_j(\xi)]$ with $A_{j\cdot}$ being j th row of matrix A . Problem 4.5 is a convex optimization problem (has an LP formulation) and provides a feasible solution to CCSP, thus $v_\epsilon^{\text{CVaR}} \geq v_\epsilon^*$. In the following we show that the approximation quality from this approach can be arbitrarily bad.

Proposition 22. *There exists instances of covering CCSP for which $v_\epsilon^{\text{CVaR}}/v_\epsilon^* = \infty$.*

Proof. Let $S \in \mathbb{R}_+$, $\Xi = \{\xi^i\}_{i \in \mathcal{N}}$, and $\xi^i = 1, i \in [\epsilon N]$ and $\xi^i = 0, i \in \mathcal{N} \setminus [\epsilon N]$. Also let $A(\xi) = 1, b(\xi) = \xi$ and the objective is to minimize x . Thus, (4.5) is equivalent to

$$v_\epsilon^{\text{CVaR}} = \min_{x \geq 0} \left\{ x : \inf_\beta [-\epsilon\beta + \epsilon(1 - x + \beta)_+ + (1 - \epsilon)(-x + \beta)_+] \leq 0 \right\},$$

while (4.1) is equivalent to

$$v_\epsilon^* = \min_{x,z} \left\{ x : x \geq z_i, i \in [\epsilon N], x \geq 0z_i, i \in \mathcal{N} \setminus [\epsilon N], z \in Z_\epsilon \cap \mathbb{B}^N \right\},$$

By simple calculation, we observe that $v_\epsilon^{\text{CVaR}} = 1$ while $v_\epsilon^* = 0$. Thus, $v_\epsilon^{\text{CVaR}}/v_\epsilon^* = \infty$. \square \square

4.2.2 Scenario approximation

The scenario approximation (SA) approach, proposed by [23], uses \bar{N} i.i.d. samples $\{\xi^i\}_{i \in [\bar{N}]}$ from the distribution \mathbb{P} and considers an optimization problem where each sampled scenario is required to be satisfied. For the covering CCSP considered here, the approximation problem is

$$v_\epsilon^{SA} = \min_x \left\{ c^\top x : x \in S, A^i x \geq b^i, \forall i \in [\bar{N}] \right\}. \quad (4.6)$$

It is shown in [23], that when the sample size \bar{N} satisfies

$$\bar{N} \geq \left\lceil \frac{2}{\epsilon} \log \left(\frac{1}{\delta} \right) + \frac{2n}{\epsilon} \log \left(\frac{2}{\epsilon} \right) + 2n \right\rceil,$$

then, with probability at least $1 - \delta$, the approximate problem (4.6) produces a feasible solution to CCSP, i.e. $v_\epsilon^{SA} \geq v_\epsilon^*$.

However, next proposition shows that with high probability, $v_\epsilon^{SA}/v_\epsilon^* = \infty$.

Proposition 23. *There exists instances that $v_\epsilon^{SA}/v_\epsilon^* = \infty$ with high probability.*

Proof. Let us consider the same instance as Proposition 22. We would like to minimize x .

Clearly, in this case, $v_\epsilon^* = 0$. Problem (4.6) is equivalent to

$$v_\epsilon^{SA} = \min_x \left\{ x : x \geq b(\xi^j), \forall j \in [\bar{N}] \right\},$$

where the sample size of $\{b(\xi^j)\}_{j \in [\bar{N}]}$ is equal to

$$\bar{N} = \left\lceil \frac{2}{\epsilon} \log \left(\frac{1}{\delta} \right) + \frac{2}{\epsilon} \log \left(\frac{2}{\epsilon} \right) + 2 \right\rceil.$$

Note that the probability that none of $\{b(\xi^j)\}_{j \in [\bar{N}]}$ is equal to 1 is bounded by

$$(1 - \epsilon)^{\bar{N}} \leq (1 - \epsilon)^{\frac{2}{\epsilon} \log(\frac{1}{\delta}) + \frac{2}{\epsilon} \log(\frac{2}{\epsilon}) + 2} \leq \delta^2 \epsilon^2,$$

where the second inequality is due to $(1 - \epsilon)^{\frac{1}{\epsilon}} \leq e^{-1}$. Thus, in this example, with probability at least $1 - \delta^2 \epsilon^2$, we can get $v_\epsilon^{SA} = 1$; i.e. $v_\epsilon^{SA}/v_\epsilon^* = \infty$. \square \square

4.2.3 Heuristic algorithm in [4]

In [4], the authors proposed a heuristic algorithm for a covering CCSP with discrete distribution which was reported to solve most of the numerical instances near-optimally. Here we show that, in general, the heuristic solution could be at least N times away from the true optimal value v_ϵ^* .

The key idea of this heuristic algorithm is to minimize the sum of infeasibilities for each scenario when the objective value is bounded:

$$\min_{x \in S, s \in \mathbb{R}_+^N} \sum_{i \in \mathcal{N}} s_i \quad (4.7a)$$

$$\text{s.t. } A^i x \geq b^i(1 - s_i), \quad i \in \mathcal{N}, \quad (4.7b)$$

$$c^\top x \leq y. \quad (4.7c)$$

The detailed procedure is described in Algorithm 4. Let v^{heur} denote the solution given by Algorithm 4.

Proposition 24. *There exists instances that $v_\epsilon^{heur}/v_\epsilon^* \geq N$.*

Proof. Let $S = \mathbb{R}_+^2$, $\xi^i = (1, 0), i \in [\epsilon N]$ and $\xi^i = (1, 1), i \in \mathcal{N} \setminus [\epsilon N]$. Also let $A(\xi) = \xi, b(\xi) = 1$ and the objective be to minimize $\frac{1}{1-\epsilon}x_1 + x_2$. Thus, (4.7) is equivalent to

$$\begin{aligned} \min_{(x,s) \in \mathbb{R}_+^{N+2}} \quad & \sum_{i \in \mathcal{N}} s_i \\ \text{s.t.} \quad & \frac{1}{1-\epsilon}x_1 + x_2 \leq y, \\ & x_1 \geq 1 - s_i, i \in [\epsilon N], \\ & x_1 + x_2 \geq 1 - s_i, i \in \mathcal{N} \setminus [\epsilon N], \end{aligned}$$

while $v_\epsilon^* = 1$.

Without loss of generality, suppose in Algorithm 4, we start with any $U > \frac{1}{1-\epsilon}$ and $L = 0$. Then, for any $y \in [1, \frac{1}{1-\epsilon})$, we must have $s_i = 1 - (1 - \epsilon)y > 0$ for each $i \in \mathcal{N}$ and $x_1 = (1 - \epsilon)y, x_2 = 0$ in the optimal solution.

Thus, in this example, we can get $v_\epsilon^{heur} = \frac{1}{1-\epsilon}, v_\epsilon^* = 1$. Thus, in the worst case, when $\epsilon = \frac{N-1}{N}$, then $v_\epsilon^{heur}/v_\epsilon^* = N$. □ □

Algorithm 4 Heuristic of [4]

```
1: Let  $L > -\infty$  and  $U < \infty$  be known lower and upper bounds for (4.1), let  $\hat{\delta} > 0$  be the
   stopping tolerance parameter.
2: while  $U - L > \hat{\delta}$  do
3:    $y \leftarrow (L + U)/2$ .
4:   Let  $(\hat{x}, \hat{s})$  be an optimal solution of (4.7) and set  $\hat{z}_i = \mathbb{I}(\hat{s}_i = 0)$  for all  $i \in \mathcal{N}$ .
5:   if  $\sum_{i=1}^N \hat{z}_i \geq N - \epsilon N$  then
6:      $U \leftarrow y$ .
7:   else
8:      $L \leftarrow y$ .
9:   end if
10: end while
11: Output  $v_\epsilon^{huer} \leftarrow U$ .
```

4.2.4 Single-criterion inapproximability

Here we show that unless $P=NP$, it is impossible to obtain a polynomial time algorithm with single-criterion approximation factor. This motivates the need for bicriteria approximations. Our reduction is similar to the one used in [41], where they showed that when $\sigma = 1$, κ -edge dense graph can be reduced to (4.1) with binary x ; however, they did not prove the inapproximability result. Here, we consider continuous covering problems as opposed to combinatorial ones, and prove the inapproximability results for both $\sigma = 1$ and $\gamma = 1$.

Theorem 16. *Suppose we have a polynomial time algorithm that returns a (σ, γ) -approximate solution to a covering CCSP with a discrete distribution with N realizations. Then*

- (i) *if $\gamma = 1$, then we must have $\sigma = 1/\epsilon - f(N)(1 - \epsilon)/\epsilon$ for some function f such that $f(N) \rightarrow 0$ as $N \rightarrow \infty$;*
- (ii) *if $\sigma = 1$, then we must have $\gamma = g(N)$ for some function g such that $g(N) \rightarrow \infty$ as $N \rightarrow \infty$.*

Proof. (i) Consider the NP-complete problem κ -dense graph which asks

(κ nodes- dense graph) Given a graph $G(V, E)$ with nodes V with $|V| = n$ and edges E with $|E| = N$, does it contain a dense subgraph with κ nodes with number of edges at least $N(1 - \epsilon)$?

This problem can be formulated as (4.2) to minimize the number of selected nodes, where $[n] = V, \mathcal{N} = E$, $x_j = 1, z_i = 1$ denote node j and edge i are chosen respectively, and $j \in \text{adj}(i)$ if edge i contains nodes j and $j \notin \text{adj}(i)$, 0 otherwise, i.e.,

$$v_\epsilon^* = \min_{x,z} \sum_{j \in [n]} x_j \quad (4.8a)$$

$$\text{s.t. } x_j \geq z_i, \forall i \in \mathcal{N}, \forall j \in \text{adj}(i), \quad (4.8b)$$

$$\sum_{i \in \mathcal{N}} z_i \geq N - \epsilon N, \quad (4.8c)$$

$$z_i \in \{0, 1\}, \forall i \in \mathcal{N}. \quad (4.8d)$$

Now suppose we get an approximate solution, i.e. a subgraph with number of nodes $v_\epsilon^* = \kappa$ and number of edges $N - \sigma\epsilon N$. By Theorem 1.2 in [6], there is no polynomial approximation algorithm for κ nodes- dense graph with constant factor, i.e.

$$\frac{N - \sigma\epsilon N}{N - \epsilon N} = f(N),$$

with some function $f(\cdot)$ such that $\lim_{N \rightarrow \infty} f(N) = 0$. Thus, $\sigma = 1/\epsilon - f(N)(1 - \epsilon)/\epsilon$.

(ii) Now consider another variant of κ - dense graph which asks

($N(1 - \epsilon)$ edges- dense graph) Given a graph with nodes V with $|V| = n$ and edges E with $|E| = N$, does it contain a dense subgraph with number of edges at least $N(1 - \epsilon)$ and number of nodes at most κ ?

This problem can be also formulated as (4.8). Now suppose we get an approximate solution, i.e. a subgraph with number of nodes $v_\epsilon^* \leq \gamma\kappa$ and number of edges at least $N - \epsilon N$.

Now we prove the following claim.

Claim: If there exists a γ (γ is a positive integer constant) approximation algorithm of $N(1 - \epsilon)$ edges- dense graph, then there exists an $\frac{1}{2.5\gamma^2}$ approximation algorithm of κ nodes- dense graph.

Proof. Suppose $G(V, E)$ has a subgraph G' with κ nodes and $N(1 - \epsilon)$ edges. Then we claim that there exists a subgraph of G' with κ/γ nodes and at least $N(1 - \epsilon)/(2.5\gamma^2)$ edges. We prove this statement by construction and let \hat{G} be the largest subgraph of G' with κ/γ nodes and ν edges. We only need to show that $\nu \geq N(1 - \epsilon)/(2.5\gamma^2)$. First, we partition the graph into γ groups $\{\mathcal{G}_i\}_{i \in [\kappa]}$, where each group has κ/γ nodes. Next, we discuss the edges within each group and among groups.

Case 1 For each group, the number of edges is at most ν by our assumption. Here, we have λ groups.

Case 2 For each pair of group $\mathcal{G}_i, \mathcal{G}_j$ with $i \neq j$, the number of edges (cuts) linking these two groups are no larger than 5ν . Here, we have $\binom{\gamma}{2}$ pairs

Hence, total number of the edges in the subgraph G' is upper bounded by $(\gamma + 5\binom{\gamma}{2})\nu \leq 2.5\gamma^2\nu$. Thus, $\nu \geq N(1 - \epsilon)/(2.5\gamma^2)$.

Now by the hypothesis, we can find a subgraph \tilde{G} with κ nodes and at least $N(1 - \epsilon)/2.5\gamma^2$ edges in polynomial time. Thus, there is a $\frac{1}{2.5\gamma^2}$ approximation algorithm of κ nodes- dense graph.

◇

By Theorem 1.2 in [6], there is no polynomial approximation algorithm for κ nodes- dense graph with constant factor. Thus, there is no polynomial approximation algorithm $N(1 - \epsilon)$ edges- dense graph, i.e., $\gamma = g(N)$ with some function $g(\cdot)$ such that $\lim_{N \rightarrow \infty} g(N) = \infty$.

□

4.3 Bicriteria approximation

In this section we first propose a scaling algorithm based on solving a continuous relaxation of the MIP formulation (4.2) of a covering CCSP. Then we provide a bicriteria approximation analysis of this algorithm, and show that the analysis is tight. We then show that for

covering CCSP with right hand side uncertainty a well known strengthening of the relaxation does not help in getting better approximation ratios. Finally, we illustrate the performance of the approximation algorithm on a portfolio optimization example.

4.3.1 The scaling algorithm

The proposed algorithm, described in Algorithm 5, is as follows. Given a violation ratio $\sigma \geq 1$ and an approximation ratio $\gamma \geq 1$ (depending on the violation ratio), we first solve a continuous relaxation of the MIP (4.2) by relaxing the integrality of variables z . Let $X_{\sigma\epsilon}$ be the set defined in (4.3) with ϵ replaced by $\sigma\epsilon$. Given an optimal solution \hat{x} of the continuous relaxation we can scale \hat{x} by a scalar $u \in [1, \gamma]$ to ensure that $u\hat{x} \in X_{\sigma\epsilon}$ because of the covering type constraints. By the nonnegativity of \hat{x} and choice of γ we can ensure that such a solution is a (σ, γ) -bicriteria approximation solution (see Section 4.3.2).

Algorithm 5 Scaling approximation algorithm.

- 1: Given σ and γ , the continuous relaxation set \hat{X}_ϵ .
 - 2: Let $\hat{x} \in \arg \min_x \{c^\top x : x \in \hat{X}_\epsilon\}$.
 - 3: Let $l = 1, u = \gamma$ and $\hat{\delta} > 0$ be a stopping tolerance parameter.
 - 4: **while** $u - l > \hat{\delta}$ **do**
 - 5: $\tau \leftarrow (l + u)/2$.
 - 6: **if** $\tau\hat{x} \in X_{\sigma\epsilon}$ **then**
 - 7: $u \leftarrow \tau$.
 - 8: **else**
 - 9: $l \leftarrow \tau$.
 - 10: **end if**
 - 11: **end while**
 - 12: Output $\bar{x} = u\hat{x}$.
-

4.3.2 Analysis

Note that the continuous relaxation of (4.2) is equivalent to the formulation below

$$v_{\epsilon,*}^{\text{rel}} = \min_{x,z} \left\{ c^\top x : x \in S, A^i x \geq b^i z_i, z \in Z_\epsilon \right\}. \quad (4.9)$$

Our approximation scheme is to scale the optimal solution to (4.9) so that the scaled solution is *nearly* feasible to (4.2).

Theorem 17. *Given violation ratio $\sigma \in [1, 1/\epsilon)$, let the approximation ratio $\gamma = \frac{1+\lfloor\sigma\epsilon N\rfloor}{1+\lfloor\sigma\epsilon N\rfloor-\epsilon N} \leq \frac{\sigma}{\sigma-1}$. Then, Algorithm 5 yields a (σ, γ) -bicriteria approximate solution for (4.1).*

Proof. Let (\hat{x}, \hat{z}) be an optimal solution of (4.9). Let set $I := \{i \in \mathcal{N} : \hat{z}_i \geq 1 - \frac{\epsilon N}{1+\lfloor\sigma\epsilon N\rfloor}\}$.

We claim that $|I| \geq N - \lfloor\sigma\epsilon N\rfloor$. For contradiction, suppose not. That is, $|I| \leq N - \lfloor\sigma\epsilon N\rfloor - 1$. Then

$$\begin{aligned} \frac{1}{N} \sum_{i \in \mathcal{N}} \hat{z}_i &= \frac{1}{N} \sum_{i \in I} \hat{z}_i + \frac{1}{N} \sum_{i \in \mathcal{N} \setminus I} \hat{z}_i < \frac{|I|}{N} + \left(1 - \frac{\epsilon N}{1 + \lfloor\sigma\epsilon N\rfloor}\right) \frac{N - |I|}{N} \\ &= 1 - \frac{\epsilon N}{1 + \lfloor\sigma\epsilon N\rfloor} + \frac{|I|}{N} \frac{\epsilon N}{1 + \lfloor\sigma\epsilon N\rfloor} \\ &\leq 1 - \frac{\epsilon N}{1 + \lfloor\sigma\epsilon N\rfloor} + \frac{N - \lfloor\sigma\epsilon N\rfloor - 1}{N} \frac{\epsilon N}{1 + \lfloor\sigma\epsilon N\rfloor} = 1 - \epsilon \end{aligned}$$

where the first strict inequality is because $\mathcal{N} \setminus I := \{i \in \mathcal{N} : \hat{z}_i < 1 - \frac{\epsilon N}{1+\lfloor\sigma\epsilon N\rfloor}\}$ is nonempty due to the hypothesis, and the second inequality is due to the definition of $|I| \leq N - \lfloor\sigma\epsilon N\rfloor - 1$. This contradicts the fact that $\hat{z} \in Z_\epsilon$.

Now let $\tilde{x} = \gamma \hat{x} := \frac{1+\lfloor\sigma\epsilon N\rfloor}{1+\lfloor\sigma\epsilon N\rfloor-\epsilon N} \hat{x}$ and $\tilde{z}_i = \min\{\lfloor\gamma \hat{z}_i\rfloor, 1\}$ for each $i \in \mathcal{N}$. We first show that $\tilde{x} \in X_{\sigma\epsilon}$. Indeed, since $\hat{x} \in S$, thus $\tilde{x} \in S$, and we know $|I| \geq N - \lfloor\sigma\epsilon N\rfloor$. Hence,

$$\mathbb{P}[\xi : a(\tilde{x}, \xi) < 1] \leq 1 - \sum_{i \in I} \frac{1}{N} \leq \frac{1}{N} \lfloor\sigma\epsilon N\rfloor \leq \sigma\epsilon,$$

where the first inequality is because according to the scaling, \tilde{x} satisfies all the scenarios in I and the last equality is due to $\sigma \geq \frac{\lfloor\sigma\epsilon N\rfloor}{\epsilon N}$. Thus, the output of Algorithm 5 \bar{x} must also be in the set $X_{\sigma\epsilon}$. Thus

$$\gamma v_\epsilon^* \geq \gamma v_\epsilon^{\text{rel}} := c^\top \tilde{x} \geq c^\top \bar{x}.$$

This implies that \bar{x} is a (γ, σ) bicriteria approximate solution for any given $\sigma \in [1, 1/\epsilon)$.

Simple calculation shows that $\frac{\sigma}{\sigma-1}$ is an upper bound of $\frac{1+\lfloor\sigma\epsilon N\rfloor}{1+\lfloor\sigma\epsilon N\rfloor-\epsilon N}$. \square

Corollary 3. *Suppose violation ratio $\sigma = 1$, then Algorithm 5 yields a single-criterion approximation for (4.1) with approximation ratio $\gamma = \epsilon N + 1$.*

Proof. The results follow by letting $\sigma = 1$ in Theorem 17 and recalling that by assumption ϵN is an integer. \square

The next proposition illustrates that the approximation ratio in Theorem 17 is tight.

Proposition 25. *Given violation ratio $\sigma \in [1, 1/\epsilon)$, there exists an instance for which Algorithm 5 can yield a solution with approximation ratio equal to $\gamma = \frac{1 + \lfloor \sigma \epsilon N \rfloor}{1 + \lfloor \sigma \epsilon N \rfloor - \epsilon N}$.*

Proof. Let $S = \{(x, w) \in \mathbb{R}_+^2 : x + w \geq 1\}$, $\Xi = \{\xi^i\}_{i \in \mathcal{N}}$, and $\xi^i = \gamma, i \in [\epsilon N]$ and $\xi^i = 1, i \in \mathcal{N} \setminus [\epsilon N]$. Also let $A(\xi) = [0, \xi]$, $b(\xi) = 1$ and the objective is to minimize $x + w$. Thus, (4.2) is equivalent to

$$v_\epsilon^* = \min_{x \geq 0, w \geq 0, z \in Z_\epsilon \cap \mathbb{B}^N} \{x + w : x + w \geq 1, \gamma x \geq z_i, i \in [1 + \lfloor \sigma \epsilon N \rfloor], \\ x \geq z_i, i \in \mathcal{N} \setminus [1 + \lfloor \sigma \epsilon N \rfloor]\}.$$

In this example, we must have $v_\epsilon^{\text{rel}} = v_\epsilon^* = 1$ and one of the optimal solution of (4.9) is $\hat{x} = \frac{1}{\gamma}, \hat{w} = 1 - \frac{1}{\gamma}, \hat{z}_i = \frac{1}{\gamma}$ for $i \in [1 + \lfloor \sigma \epsilon N \rfloor]$ and 1, otherwise. In this case,

$$\sum_{i \in [N]} \min\{\lfloor \tau \hat{z}_i \rfloor, 1\} \geq N - \lfloor \sigma \epsilon N \rfloor$$

if and only if $\tau \geq \gamma$. Thus the solution returned by Algorithm 5 has an optimal approximation ratio of γ . □ □

4.3.3 Effect of tighter relaxation

In this subsection, we study a special class of discrete CCSPs (i.e., CCSPs with right-hand side uncertainty) where we assume that $A^i = A$ for each $i \in \mathcal{N}$. Thus, in this case, (4.2) can be reformulated as

$$v_\epsilon^* = \min_{x, y, z} \left\{ c^\top x : x \in S, Ax \geq y, y \geq b^i z_i, z \in Z_\epsilon \cap \mathbb{B}^N \right\}. \quad (4.10)$$

Recently, many researchers tried to develop valid inequalities for the following cardinality constrained mixing set:

$$Y := \text{conv} \left\{ (y, z) : y \geq b^i z_i, \forall i \in \mathcal{N}, z \in Z_\epsilon \cap \mathbb{B}^N \right\}, \quad (4.11)$$

see [1, 56, 66, 68] for example, while the complete description of Y has not been discovered yet. It is well known that in general case, the separation over the set Y is NP-hard.

However, some computational study shows that some classes of valid inequalities of (4.11) help speedup the solution algorithms.

Now suppose that Y is known and consider the following relaxation of (4.10) as

$$v_{\epsilon,R}^{\text{rel}} = \min_{x,z} \left\{ c^\top x : x \in S, Ax \geq y, (y,z) \in Y \right\}. \quad (4.12)$$

The following theorem shows that in the worst case, adding all the valid inequalities from the mixing set still yields the same approximation ratio as that in Theorem 17.

Proposition 26. *Given a violation ratio $\sigma \in [1, 1/\epsilon)$. Let \widehat{X}_ϵ be defined by the feasible region of (4.12). Then*

- (i) *Algorithm 5 yields a (σ, γ) -bicriteria approximate solution for (4.1), where $\gamma = \frac{1 + \lfloor \sigma \epsilon N \rfloor}{1 + \lfloor \sigma \epsilon N \rfloor - \epsilon N}$;*
- (ii) *the approximation ratio is tight.*

Proof. The statement in Part (i) follows directly from Theorem 17 since (4.12) is at least as strong a relaxation as (4.9). Thus the approximation guarantees hold.

Part (ii): For any given violation ratio $\sigma \in [1, 1/\epsilon)$ and $\gamma = \frac{1 + \lfloor \sigma \epsilon N \rfloor}{1 + \lfloor \sigma \epsilon N \rfloor - \epsilon N}$, let us consider the same example used in Proposition 25, where $S = \{(x, w) \in \mathbb{R}_+^2 : x + w \geq 1\}$, and each scenario $X^i = \{x \in \mathbb{R}_+ : x \geq \frac{1}{\gamma}\}$ for $i \in [1 + \lfloor \sigma \epsilon N \rfloor]$, while $X^i = \{x \in \mathbb{R}_+ : x \geq 1\}$ for $i \in \mathcal{N} \setminus [1 + \lfloor \sigma \epsilon N \rfloor]$. Thus,

$$v_{\epsilon,R}^{\text{rel}} = \min_{x,w,z} \{x + w : x + w \geq 1, x \geq y_i, i \in \mathcal{N}, (y, z) \in Y\},$$

and

$$Y = \text{conv} \left\{ (y, z) : \begin{array}{ll} y_i \geq \frac{1}{\gamma} z_i, & \forall i \in [1 + \lfloor \sigma \epsilon N \rfloor] \\ y_i \geq z_i, & \forall i \in \mathcal{N} \setminus [1 + \lfloor \sigma \epsilon N \rfloor] \\ z \in Z_\epsilon \cap \mathbb{B}^N \end{array} \right\}.$$

We only need to show that the above relaxation is equivalent to the continuous relaxation

(4.9). Recall that

$$\bar{Y} = \left\{ (y, z) : \begin{array}{l} y_i \geq \frac{1}{\gamma} z_i, \forall i \in [1 + \lfloor \sigma \epsilon N \rfloor] \\ y_i \geq z_i, \forall i \in \mathcal{N} \setminus [1 + \lfloor \sigma \epsilon N \rfloor] \\ \sum_{i \in \mathcal{N}} z_i \geq N - \epsilon N, z \in [0, 1]^N \end{array} \right\}.$$

We need to show that

Claim: $Y = \bar{Y}$.

Proof of Claim: We know that $Y \subseteq \bar{Y}$. Thus, it remains to show that $Y \supseteq \bar{Y}$. Or equivalently, we can show that all the extreme points of \bar{Y} belong to Y . Indeed, given an extreme point $(\hat{y}, \hat{z}) \in \mathbb{R}^N \times \mathbb{R}^N$, we know that it should satisfy exactly $2N$ equalities. Note that in the set \bar{Y} , except the bounds of z variables (i.e., $z \in [0, 1]^N$), there are only $N + 1$ additional inequalities, which could be binding. Thus at least $(N - 1)$ z variables of extreme points (\hat{y}, \hat{z}) should satisfy their bound constraints. Now let $\mathcal{T} = \{i \in \mathcal{N} : \hat{z}_i \in \mathbb{B}\}$, and we have $|\mathcal{T}| \geq N - 1$. Suppose that $|\mathcal{T}| = N - 1$. Now there are two cases:

Case 1. $\sum_{i \in \mathcal{N}} \hat{z}_i = N - \epsilon N$. Then since $|\mathcal{T}| = N - 1$, $\sum_{i \in \mathcal{T}} \hat{z}_i + \sum_{i \in \mathcal{N} \setminus \mathcal{T}} \hat{z}_i = N - \epsilon N$ and ϵN is a nonnegative integer, thus clearly, $\hat{z}_i \in \mathbb{B}$ for all $i \in \mathcal{N}$, contradiction.

Case 2. $\sum_{i \in \mathcal{N}} \hat{z}_i > N - \epsilon N$. Then in the set \bar{Y} , except the boundaries of z variables (i.e., $z \in [0, 1]^N$), there are only N additional inequalities, which could be binding. Then, clearly, $\hat{z}_i \in \mathbb{B}$ for all $i \in \mathcal{N}$, contradiction.

◇

By the claim above, we know that in this example, (4.12) is equivalent to (4.9). Hence, by the tightness of approximation ratio in Proposition 25, we know that mixing set inequalities cannot improve the performance of scaling Algorithm 5. □

4.3.4 Numerical illustration

In this subsection, we present a numerical illustration of the proposed approximation algorithm. We consider the following chance constrained portfolio optimization problem studied

in [76, 82], which is to minimize the investment cost under specified return level; i.e.

$$v_\epsilon^* = \min_{x,z} \left\{ c^\top x : x \geq 0, \mathbb{P}[\boldsymbol{\xi}^\top x < 1] \leq \epsilon \right\}, \quad (4.13)$$

where $\boldsymbol{\xi} \in \mathbb{R}_+^n$ represents the random return with realizations $\{\xi^i\}_{i \in [N]}$. We use the data from [82], where $n = 50$ and $\boldsymbol{\xi}$ has finite support with $N = 100$ scenarios and risk parameter $\epsilon \in \{0.05, 0.10\}$.

The computational results are shown in Figure 5. For any violation ratio $\sigma \in [1, 1/\epsilon)$, we let \hat{v}_σ denote the output objective value of Algorithm 5. In Figure 5, the blue starred curve denotes the theoretical approximation ratio γ proposed in Theorem 4, and red squared curve denoted as $\frac{\hat{v}_\sigma}{v_{\epsilon,*}^{\text{rel}}}$ is the upper bound of the practical approximation ratio $\frac{\hat{v}_\sigma}{v_\epsilon^*}$.

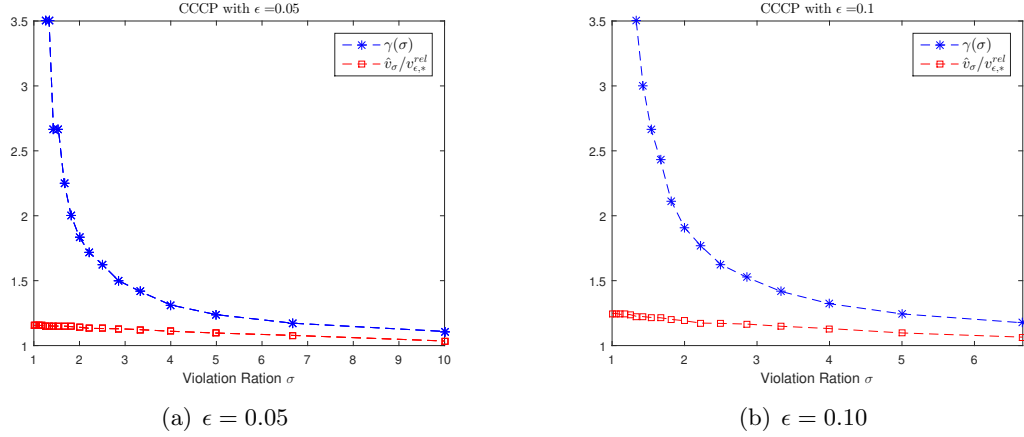


Figure 5: Illustration of approximation ratios for a CCSP with discrete support.

In Figure 5, we see that when violation ratio σ is close to 1, the theoretical ratio (starred curve) is quite large – almost equal to $\epsilon N + 1$. However, the actual scaling solution is around 1.3 away from the true optimal for both $\epsilon = 0.05$ and 0.1. When σ increases, the theoretical ratio (starred curve) decreases dramatically. Typically, if we choose $\sigma = 2$, then the theoretical approximation ratio (starred curve) is around 2 but the practical ratio (squared curve) is around 1.2. However, when $\sigma > 2$, the practical approximation ratio does not improve too much. Thus, for these instances, we suggest choosing $\sigma \in [1, 2]$.

4.4 *Extension: approximation algorithm for a CCSP with general probability mass functions*

In this section, assume that each scenario ξ^i corresponds to a probability mass $p_i \in \mathbb{Q}_+$ for $i \in \mathcal{N}$. Under this assumption, similarly by introducing binary variables z_i for $i \in [N]$, then we can reformulate (4.1) as the mixed integer program (MIP) (c.f. [82, 120]) as

$$v_\epsilon^* = \min_{x,z} \left\{ c^\top x : x \in S, A^i x \geq b^i z_i, z \in Z_\epsilon \cap \mathbb{B}^N \right\}, \quad (4.14)$$

where

$$Z_\epsilon := \left\{ z \in [0, 1]^N : \sum_{i \in [N]} p_i z_i \geq 1 - \epsilon \right\}, \mathbb{B} = \{0, 1\}.$$

Since $\{p_i\}_{i \in \mathcal{N}} \subseteq \mathbb{Q}$, let q be the greatest common divisor of the denominators of $\{p_i\}_{i \in \mathcal{N}}$. Let us round $\sigma\epsilon$ for any $\sigma \in [1, 1/\epsilon)$ below:

$$k_{\epsilon,q}(\sigma) := \lfloor \sigma\epsilon q \rfloor / q, \quad (4.15)$$

which is no larger than ϵ/r . We remark that if $\{p_i\}_{i \in \mathcal{N}}$ are identical, then $q = N$ and $k_{\epsilon,N}(r) = \lfloor \sigma\epsilon N \rfloor / N$.

By definition of $k_{\epsilon,q}(\sigma)$ and observing that in (4.16), $z \in \{0, 1\}^N$ is integral, we could replace set Z_ϵ by $Z_{k_{\epsilon,q}(1)} \cap \{0, 1\}^N$ in (4.16), i.e., (4.16) is equivalent to

$$v_\epsilon^* = \min_{x,z} \left\{ c^\top x : x \in S, A^i x \geq b^i z_i, z \in Z_{k_{\epsilon,q}(1)} \cap \mathbb{B}^N \right\}, \quad (4.16)$$

with $k_{\epsilon,q}(1) = \lfloor q\epsilon \rfloor / q$.

We now study the approximation results of the CCSP with finite support. Recall that $\phi(s) = \min(s, 1)$, hence, the continuous relaxation of (4.16) is equivalent to the formulation below

$$v_{\epsilon,*}^{\text{rel}} = \min_{x,z} \left\{ c^\top x : x \in S, A^i x \geq b^i z_i, z \in Z_{k_{\epsilon,q}(1)} \right\}, \quad (4.17)$$

which turns out to be continuous relaxation of (4.16).

Our approximation scheme is to scale the optimal solution to (4.17).

Theorem 18. *Given violation ratio $\sigma \in [1, 1/\epsilon)$, define*

$$\alpha_{\epsilon,q}(\sigma) := \min_{S \subseteq \mathcal{N}} \left\{ 1 - k_{\epsilon,q}(\sigma) - \sum_{i \in S} p_i : \sum_{i \in S} p_i < k_{\epsilon,q}(\sigma) \right\}. \quad (4.18)$$

In Algorithm 5, let approximation ratio $\gamma = \frac{\alpha_{\epsilon,q}(\sigma) + k_{\epsilon,q}(\sigma)}{\alpha_{\epsilon,q}(\sigma) + k_{\epsilon,q}(\sigma) - k_{\epsilon,q}(1)}$, and the relaxed set \widehat{X}_ϵ be the feasible region of (4.17). Then, Algorithm 5 yields (γ, σ) bicriteria approximation guarantee for (4.1).

Proof. First of all, from the discussion above, we note that the risk parameter in (4.1) can be replace by $k_{\epsilon,q}(1)$.

Let $(\widehat{x}, \widehat{z})$ be an optimal solution of (4.17). Let set $I := \{i \in \mathcal{N} : \widehat{z}_i \geq 1 - \frac{k_{\epsilon,q}(1)}{\alpha_{\epsilon,q}(\sigma) + k_{\epsilon,q}(\sigma)}\}$. Now we make the following claim.

Claim: $\sum_{i \in I} p_i \geq 1 - k_{\epsilon,q}(\sigma)$.

Proof. Suppose not. Thus, $\sum_{i \in I} p_i < 1 - k_{\epsilon,q}(\sigma)$. Then

$$\begin{aligned} \sum_{i \in \mathcal{N}} p_i \widehat{z}_i &= \sum_{i \in I} p_i \widehat{z}_i + \sum_{i \in \mathcal{N} \setminus I} p_i \widehat{z}_i < \sum_{i \in I} p_i + \left(1 - \frac{k_{\epsilon,q}(1)}{\alpha_{\epsilon,q}(\sigma) + k_{\epsilon,q}(\sigma)}\right) \sum_{i \in \mathcal{N} \setminus I} p_i \\ &= 1 - \frac{k_{\epsilon,q}(1)}{\alpha_{\epsilon,q}(\sigma) + k_{\epsilon,q}(\sigma)} + \frac{k_{\epsilon,q}(1)}{\alpha_{\epsilon,q}(\sigma) + k_{\epsilon,q}(\sigma)} \sum_{i \in I} p_i \\ &\leq 1 - \frac{k_{\epsilon,q}(1)}{\alpha_{\epsilon,q}(\sigma) + k_{\epsilon,q}(\sigma)} + \frac{k_{\epsilon,q}(1)}{\alpha_{\epsilon,q}(\sigma) + k_{\epsilon,q}(\sigma)} [1 - k_{\epsilon,q}(\sigma) - \alpha_{\epsilon,q}(\sigma)] \\ &= 1 - k_{\epsilon,q}(1) \end{aligned}$$

where the first strict inequality is because $\mathcal{N} \setminus I := \{i \in \mathcal{N} : \widehat{z}_i < 1 - \frac{k_{\epsilon,q}(1)}{\alpha_{\epsilon,q}(\sigma) + k_{\epsilon,q}(\sigma)}\}$ is nonempty due to the hypothesis, and the second inequality is due to the definition of $\alpha_{\epsilon,q}(\sigma)$. This contradicts the fact that $\widehat{z} \in \widehat{Z}$. \diamond

Now let $\tilde{x} = \gamma \widehat{x} := \frac{\alpha_{\epsilon,q}(\sigma) + k_{\epsilon,q}(\sigma)}{\alpha_{\epsilon,q}(\sigma) + k_{\epsilon,q}(\sigma) - k_{\epsilon,q}(1)} \widehat{x}$ and $\tilde{z}_i = \min\{\lfloor \gamma \widehat{z}_i \rfloor, 1\}$ for each $i \in \mathcal{N}$. We first show that $\tilde{x} \in X_{k_{\epsilon,q}(\sigma)}$. Indeed, since $\widehat{x} \in S$, thus $\tilde{x} \in S$, and by Claim, we know $\sum_{i \in I} p_i \geq 1 - k_{\epsilon,q}(\sigma)$. Hence,

$$\mathbb{P}[\xi : a(\tilde{x}, \xi) < 1] \leq 1 - \sum_{i \in I} p_i \leq k_{\epsilon,q}(\sigma).$$

Thus, the output of Algorithm 5 \bar{x} must also be in set $X_{k_{\epsilon,q}(\sigma)}$.

By the description of Algorithm 5, we have

$$\gamma v_\epsilon^* \geq \gamma v_{\epsilon,\phi}^{\text{rel}} := c^\top \tilde{x} \geq c^\top \bar{x}.$$

This implies that \bar{x} yields (γ, σ) bicriteria approximate solution. \square

One direct application of Theorem 18 is the single-criterion approximation.

Corollary 4. *Let q be the greatest common divisor of the denominators in $\{p_i\}_{i \in \mathcal{N}}$ and violation ratio $\sigma = 1$, then Algorithm 5 yields a single-criterion approximation for (4.1) with approximation ratio $\gamma := \frac{k_{\epsilon,q}(1)}{\alpha_{\epsilon,q}(1)} + 1$, where $k_{\epsilon,q}(1), \alpha_{\epsilon,q}(1)$ are defined in (4.15), (4.18), respectively.*

Proof. The results follow directly from Theorem 18 by letting $\sigma = 1$. \square

CHAPTER V

ON DETERMINISTIC REFORMULATIONS OF DISTRIBUTIONALLY ROBUST JOINT CHANCE CONSTRAINED OPTIMIZATION PROBLEMS

5.1 Introduction

5.1.1 Problem Setting

We consider a distributionally robust chance constrained program (DRCCP) of the form (c.f. [24, 46, 51, 123]):

$$v^* = \min c^\top x, \tag{5.1a}$$

$$\text{s.t. } x \in S, \tag{5.1b}$$

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\xi : F(x, \xi) \geq 0] \geq 1 - \epsilon. \tag{5.1c}$$

where $x \in \mathbb{R}^n$ is a decision vector; the vector $c \in \mathbb{R}^n$ denotes the objective coefficients; the set $S \subseteq \mathbb{R}^n$ denotes deterministic constraints on x ; the random vector ξ supported on $\Xi \subset \mathbb{R}^m$ denotes uncertain constraint coefficients; the mapping $F(x, \xi) := (f_1(x, \xi), \dots, f_I(x, \xi))^\top$ with $f_i(x, \xi) : \mathbb{R}^n \times \Xi \rightarrow \mathbb{R}$ for all $i \in [I] := \{1, \dots, I\}$ defines a set of uncertain constraints on x ; the ambiguity set \mathcal{P} denotes a set of probability measures \mathbb{P} on the space Ξ with a sigma algebra \mathcal{F} ; and $\epsilon \in (0, 1)$ denotes a risk tolerance. In (5.1) we seek a decision vector x to minimize a linear objective $c^\top x$ subject to a set of deterministic constraints defined by S , and a chance constraint (5.1c) that is required to hold for any probability distribution from the ambiguity set \mathcal{P} with a probability of $1 - \epsilon$. Note that when $|I| = 1$ the constraint (5.1c) involves a *single* chance constraint and if $|I| \geq 2$ it involves a *joint* chance constraint.

The primary difficulty of (5.1) is due to the distributionally robust chance constraint (5.1c). Let us denote the feasible region induced by (5.1c) as

$$Z := \left\{ x \in \mathbb{R}^n : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\xi : F(x, \xi) \geq 0] \geq 1 - \epsilon \right\}. \tag{5.2}$$

In this chapter we study deterministic reformulations of the set Z and its convexity properties. Our study is restricted to the convex, moment constrained setting (cf. [93, 95]), i.e. under the following assumptions.

- (A1) Each function $f_i(x, \xi)$ in the mapping $F(x, \xi) := (f_1(x, \xi), \dots, f_I(x, \xi))^\top$ is concave in x for any fixed ξ , and is convex in ξ for any fixed x .
- (A2) The random vector ξ is supported on a nonempty closed convex set $\Xi \subseteq \mathbb{R}^m$.
- (A3) The ambiguity set \mathcal{P} is nonempty and is defined by moment constraints:

$$\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\phi_t(\xi)] = g_t, t \in \mathcal{T}_1, \mathbb{E}_{\mathbb{P}}[\phi_t(\xi)] \geq g_t, t \in \mathcal{T}_2, \} \quad (5.3)$$

where $\mathcal{P}_0(\Xi)$ denotes the set of all of probability measures on Ξ with a sigma algebra \mathcal{F} , and for each $t \in \mathcal{T}_1 \cup \mathcal{T}_2$, the moment function $\phi_t : \Xi \rightarrow \mathbb{R}$ is a real valued continuous function and g_t is a scalar. Furthermore, for each $t \in \mathcal{T}_1$, the function $\phi_t(\xi)$ is linear, and for each $t \in \mathcal{T}_2$, the function $\phi_t(\xi)$ is concave.

5.1.2 Contributions

Even under the above convexity assumptions the set Z is nonconvex in general, making (5.1) a difficult optimization problem. Moreover it is not described by explicit functions, and so is not suitable for direct optimization as a mathematical program. In this chapter we first provide a deterministic approximation of Z that is nearly tight and then identify a variety of settings under which Z is convex. The main results of this chapter are summarized next.

1. We propose a deterministic conservative approximation with its closure equal to Z , which is in general nonconvex and can be formulated as an optimization problem involving biconvex constraints.
2. If there is a single uncertain constraint, i.e. $|I| = 1$, we prove that the proposed deterministic approximation is exact and reduces to a tractable convex program. This result is a generalization to existing works (e.g., [24, 123, 135]) with arbitrary convex ambiguity set rather than known first and second moments.

3. We prove that if the ambiguity set \mathcal{P} contains only one moment inequality, i.e. $|\mathcal{T}_1| = 0$ and $|\mathcal{T}_2| = 1$, then Z is a tractable convex program; and if the ambiguity set contains only one moment linear equality, i.e. $|\mathcal{T}_1| = 1$ and $|\mathcal{T}_2| = 0$, then Z is equivalent to the disjunction of two tractable convex programs.
4. We prove that if $\Xi = \mathbb{R}^m$ and the moment functions $\{\phi_t(\xi)\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$ are linear, then Z is equivalent to the feasible region of a robust convex program.
5. We prove that if Ξ is a closed convex cone, the function $f_i(x, \xi)$ for any $i \in [I]$ is of the (separable) form $f_i(x, \xi) = w_i(x) - h_i(\xi)$ where $h_i(\xi)$ is positively homogeneous on Ξ , and the moment functions $\{\phi_t(\xi)\}_{t \in \mathcal{T}_2}$ are positively homogeneous on Ξ , then set Z is convex. This result is a generalization of [47], where the authors assumed that $w_i(\cdot), h_i(\cdot)$ are affine functions for each $i \in [I]$.
6. When the decision variables are pure binary (i.e. $S \subseteq \{0, 1\}^n$) and uncertain constraints are linear, we show that the proposed deterministic approximation can be reformulated as a mixed integer convex program. We also present a numerical study to demonstrate that the proposed reformulation can be effectively solved using a standard solver.

5.1.3 Connection to existing works

Recently, nonlinear uncertain constraints have been extensively studied in different areas, e.g., wireless communication [55], transportation [123], facility location [58], power system [124] and so on. Many researchers investigated distributionally robust program with various nonlinear moment ambiguity sets, for example, [47] studied mean dispersion ambiguity set, [30, 33, 51, 135] incorporated second moment into ambiguity set, and [116] considered coefficient of variation. Next, we will review single and joint chance constraints separately.

In the case of a single uncertain constraint, i.e., $|I| = 1$, there has been significant efforts in identifying settings where Z can be reformulated by deterministic convex constraints. For example, with known mean and covariance of ξ , the authors in [24] showed that the set Z can be formulated as a second order cone program (SOCP). Recently, more efforts

have been made to derive tractable reformulation of the set Z . For instance, in [135], the authors showed that with given range of first- and second- order moments, the set Z can be reformulated as a semidefinite program (SDP). These tractability results have been generalized to nonlinear uncertain constraints in [123]. In [46], the authors demonstrated that the set Z is convex when \mathcal{P} involves conic moment constraints or unimodality of \mathbb{P} . Generalizing the above mentioned earlier works, this chapter demonstrates that for any ambiguity set with convex moment constraints, when there is a single chance constraint, the set Z can be reformulated as a convex program.

Tractability results for a joint DRCCP (i.e., $|I| > 1$) are very rare. It has been shown in [46] that optimization over the set Z is in general NP-hard. Therefore, much of the earlier works built approximation of the set Z instead of deriving its exact reformulation. For example, in [73], the authors suggested that using Bonferroni's inequality to decompose a joint chance constraint into $|I|$ different single chance constraints whose sum of risk parameters is no larger than ϵ . With such decomposition, any approximation scheme proposed for a single chance constraint could be directly applied. However, Bonferroni's inequality is not tight in general (c.f. [30, 135]). Thus, in [30], the authors proposed to improve Bonferroni's inequality by scaling each uncertain constraint with a positive number and converting them into a single constraint. For any fixed scaler, they were able to provide a conservative SOCP approximation. Later, it was shown in [135] that by optimizing over the scaling parameters, the feasible region of the proposed scaling method is nearly exact to set Z when \mathcal{P} is described by first- and second- order moments. This result was established using strong duality of SDP. However, in this case, the corresponding deterministic reformulation of (5.1) turns out to be a bilinear optimization problem, which is naturally hard to solve (c.f. [13]). Recently, [47] derived a tractable reformulation under the restricted assumption that the stochastic mapping $F(x, \xi)$ is separate and affine in (x, ξ) , Ξ is a closed convex solid cone and the ambiguity set is defined by mean and dispersion constraints, where the dispersion function is positively homogeneous on the cone Ξ . We extend the results of [135] to any ambiguity set with convex moment constraints and show that the approximation yields a mixed integer convex program when the decision vector x is binary. Unlike [47], we show

that a DRCCP with single moment constraint is tractable by relaxing their assumptions on the set Ξ and mapping $F(x, \xi)$, and we also provide new sufficient conditions under which joint DRCCP is tractable.

The remainder of the chapter is organized as follows. Section 5.2 presents some preliminary results to be used subsequently. Section 5.3 proposes an equivalent deterministic reformulation of the set Z and develops a tight approximation of the set Z via a system of biconvex constraints. Section 5.4 provides various sufficient conditions for the convexity of the set Z . Section 5.5 demonstrates that the proposed tight approximation of Z yields a mixed integer convex program when the decision variables are binary and the uncertain constraints are linear. A numerical study is presented to test the proposed formulation.

5.2 Preliminaries

We first present notations, some standard results and then define a special function associated with the set Z that will be used in our analysis.

5.2.1 Notations

We use bold ξ to denote random vector while ξ is a realization of ξ , and \mathbf{e} be all-one vector. For a positive integer m , let $[m] := \{1, \dots, m\}$ and $\mathbb{R}_+^m = \{x \in \mathbb{R}^m : x \geq 0\}$, $\mathbb{R}_{++}^m = \{x \in \mathbb{R}^m : x_i > 0, \forall i \in [m]\}$. We let $\mathcal{M}_+(\Xi)$ denote the cone of all nonnegative measures on Ξ . Given a vector $\hat{F} = (\hat{f}_1, \dots, \hat{f}_I)^\top$, the indicator function $\mathbb{I}_+(\hat{F})$ is equal to 1 if $\hat{f}_i \geq 0$ for all $i \in [I]$, 0 otherwise. For logic expression χ , we let $\mathbb{I}(\chi)$ be 1 if χ is true, 0, otherwise. Given a function $\hat{f}(\cdot)$, we use $\text{dom } \hat{f}$ to denote its domain. Given a closed convex cone \mathcal{C} , a function $f : \mathcal{C} \rightarrow \mathbb{R}$ is *positively homogeneous* on \mathcal{C} if $f(\lambda x) = \lambda f(x)$ for any $x \in \mathcal{C}$ and $\lambda \geq 0$.

5.2.2 Some standard results

As is common in the distributionally robust optimization literature (e.g., [93]), our deterministic reformulation of Z relies on dualizing the optimization problem appearing in the left-hand-side of the chance constraint defining (5.2). Towards this, in addition to Assumptions (A1) - (A3), we will make the following constraint qualification assumption on \mathcal{P}

throughout the rest of this chapter.

(A4) (Slater's condition) there exists a probability measure \mathbb{P} satisfying the constraints defining \mathcal{P} for any sufficiently small perturbation of $\{g_t\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$.

We will use the following strong duality result.

Lemma 6. *Let \mathcal{P} be defined as in (5.3) and suppose Assumptions (A1) - (A4) hold. Let $\psi(\xi)$ be a Lebesgue measurable function on (Ξ, \mathcal{F}) such that $|\mathbb{E}_{\mathbb{P}}[\psi(\xi)]| < \infty$ for any $\mathbb{P} \in \mathcal{P}$, then*

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[\psi(\xi)], \quad (5.4)$$

is equivalent to the following mathematical program:

$$\max \quad \lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t, \quad (5.5a)$$

$$\text{s.t.} \quad \lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t \leq \psi(\xi), \forall \xi \in \Xi, \quad (5.5b)$$

$$\gamma_t \geq 0, t \in \mathcal{T}_2. \quad (5.5c)$$

Proof. Note that $\mathcal{M}_+(\Xi)$ denote the cone of all nonnegative measures on Ξ . Then (5.4) can be formulated as

$$\begin{aligned} \inf_{\mu \in \mathcal{M}_+(\Xi)} \quad & \int_{\Xi} \psi(\xi) d\mu(\xi) \\ \text{s.t.} \quad & \int_{\Xi} \phi_t(\xi) d\mu(\xi) = g_t, \forall t \in \mathcal{T}_1, \int_{\Xi} \phi_t(\xi) d\mu(\xi) \geq g_t, \forall t \in \mathcal{T}_2, \int_{\Xi} d\mu(\xi) = 1. \end{aligned}$$

The dual of the above semi-infinite linear program is (5.5). Due to Assumption (A4), Theorem 5.99 in [21] implies that strong duality holds and the set of optimal solutions are bounded. \square

Next, we note that Z is closed.

Lemma 7. *Under Assumptions (A1) - (A3), Z is closed.*

Proof. For any given $\mathbb{P} \in \mathcal{P}$, let

$$Z_{\mathbb{P}} := \{x \in \mathbb{R}^n : \mathbb{P}[\xi : F(x, \xi) \geq 0] \geq 1 - \epsilon\}.$$

From Proposition 1.7 in [52], since $F(x, \xi)$ is continuous in x , $Z_{\mathbb{P}}$ is a closed set. By definition, $Z = \bigcap_{\mathbb{P} \in \mathcal{P}} Z_{\mathbb{P}}$ and it is well known that any intersection of closed set is also closed. Thus, Z is closed. \square

Finally, we mention a result from convex programming that will be useful.

Lemma 8. (*Convex Theorem of Alternatives, [12]*) *Consider the convex inequality system*

$$\begin{aligned} & f(x) < c, \\ (S_1) \quad & g_i(x) \leq 0, i \in [m], \\ & x \in X, \end{aligned}$$

where $c \in \mathbb{R}$ is a constant, $f(x), \{g_i(x)\}_{i \in [m]}$ are convex functions defined on \mathbb{R}^n and $X \subseteq \mathbb{R}^n$ is a nonempty convex set. Assume that there exist \bar{x} such that $g_i(\bar{x}) < 0$ for each $i \in [m]$. Then system (S_1) is unsolvable if and only if system (S_2) is solvable, where (S_2) is defined as

$$\begin{aligned} (S_2) \quad & \inf_{x \in X} [f(x) + \sum_{i \in [m]} \lambda_i g_i(x)] \geq c, \\ & \lambda_i \geq 0, i \in [m]. \end{aligned}$$

5.2.3 ϕ -conjugate functions

Our subsequent constructions will make use of the following function associated with the ingredients defining Z .

Definition 8. Let $f(x, \xi)$ be a function which is convex in $\xi \in \Xi$ for all x . Given functions $\{\phi_t(\xi)\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$ as defined in (5.3) the ϕ -conjugate of f corresponding to weights $(\gamma, \alpha) \in (\mathbb{R}^{|\mathcal{T}_1|} \times \mathbb{R}_+^{|\mathcal{T}_2|}) \times \mathbb{R}_+$ is

$$\psi_f(\gamma, \alpha, x) := \sup_{\xi \in \Xi} \left(\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t - \alpha f(x, \xi) \right)$$

For notational simplicity, when $f(x, \xi) = 0$ for all (x, ξ) , we denote

$$\psi_0(\gamma) := \psi_0(\gamma, \alpha, x).$$

Note that evaluating $\psi_f(\gamma, \alpha, x)$ amounts to solving a concave maximization problem. Often, such a problem is tractable. Next we establish some properties of ψ_f .

Lemma 9. *Let $f(x, \xi)$ be concave in $x \in \mathbb{R}^n$ and convex in $\xi \in \Xi$. Then its ϕ -conjugate function $\psi_f(\gamma, \alpha, x)$ has the following properties:*

- (i) $\psi_f(\cdot, \cdot, x)$ is jointly convex in $(\gamma, \alpha) \in (\mathbb{R}^{|\mathcal{T}_1|} \times \mathbb{R}_+^{|\mathcal{T}_2|}) \times \mathbb{R}_+$ for any given x ;
- (ii) $\psi_f(\cdot, \alpha, \cdot)$ is jointly convex in $(\gamma, x) \in (\mathbb{R}^{|\mathcal{T}_1|} \times \mathbb{R}_+^{|\mathcal{T}_2|}) \times \mathbb{R}^n$ for any given α ;
- (iii) if Ξ is a closed convex cone, $f(x, \cdot)$ and $\{\phi_t(\cdot)\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$ are positively homogeneous on Ξ , then

$$\psi_f(\gamma, \alpha, x) = \begin{cases} 0, & \text{if } (\gamma, \alpha, x) \in \text{dom } \psi_f, \\ +\infty, & \text{otherwise.} \end{cases}$$

Proof. Parts (i) and (ii) follow from the fact that the supremum of a set of convex functions is convex. We only show part (iii). If Ξ is a closed convex cone, $f(x, \cdot)$ and $\{\phi_t(\cdot)\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$ are positively homogeneous on Ξ , then we must have $\psi_f(\gamma, \alpha, x) \leq 0$ for any $(\gamma, \alpha, x) \in \text{dom } \psi_f$. Otherwise, there would exist $\bar{\xi} \in \Xi$ such that

$$\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\bar{\xi})\gamma_t - \alpha f(x, \bar{\xi}) > 0$$

and so

$$\sup_{\xi \in \Xi} \left(\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi)\gamma_t - \alpha f(x, \xi) \right) \geq \lim_{\lambda \rightarrow \infty} \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\lambda \bar{\xi})\gamma_t - \alpha f(x, \lambda \bar{\xi}) = \infty$$

where the first inequality comes from $\lambda \bar{\xi} \in \Xi$ for any $\lambda \geq 0$ and the first equality due to the positive homogeneity of $f(x, \cdot)$ and $\{\phi_t(\cdot)\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$. Also from positive homogeneity it follows that $\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(0)\gamma_t - \alpha f(x, 0) = 0$. Thus $\psi_f(\gamma, \alpha, x) = 0$ for all $(\gamma, \alpha, x) \in \text{dom } \psi_f$. \square

Note that in part (iii) of Lemma 9, if Ξ is a polyhedral cone, $f(x, \xi)$ is linear in ξ for any given x and $\{\phi_t(\xi)\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$ are all linear functions, then by strong duality of linear programming, $\psi_f(\gamma, \alpha, x)$ is equal to the characteristic function of a set defined by linear inequalities on (γ, α) for any given x .

Next we present a few examples of \mathcal{P} whose associated ϕ -conjugate function and its domain can be explicitly computed. We omit the calculations for brevity.

Proposition 27. *Let*

$$\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^n) : \mathbb{E}_{\mathbb{P}}[\|\xi\|_q] \leq g_1\},$$

with $q \geq 1, \Xi = \mathbb{R}^m$, and $f(x, \xi) = b - \mu^\top x - \xi^\top x$. Here $\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t = -\gamma \|\xi\|_q$. Then

$$\psi_f(\gamma, \alpha, x) = \begin{cases} -\alpha(b - \mu^\top x), & \text{if } \|\alpha x\|_{\frac{q}{q-1}} \leq \gamma, \gamma \geq 0 \\ +\infty, & \text{otherwise.} \end{cases}$$

Proposition 28. *Let*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^m) : \mathbb{E}_{\mathbb{P}}[w_t^\top \xi] = g_t, t \in \mathcal{T}_1, \mathbb{E}_{\mathbb{P}}[w_t^\top \xi] \geq g_t, t \in \mathcal{T}_2 \right\},$$

with $\Xi = \mathbb{R}^m$, $f(x, \xi) = (Ax + a)^\top \xi + Bx + b$, and $\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t = \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \gamma_t w_t^\top \xi$. Then

$$\psi_f(\gamma, \alpha, x) = \begin{cases} -\alpha(Bx + b), & \text{if } \alpha(Ax + a) = \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \gamma_t w_t, \gamma_t \geq 0 \forall t \in \mathcal{T}_2 \\ +\infty, & \text{otherwise.} \end{cases}$$

Proposition 29. *Let*

$$\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^m) : \mathbb{E}_{\mathbb{P}}[\xi] = 0, \mathbb{E}_{\mathbb{P}}[\|\xi\|_q] \leq g\},$$

with $q \geq 1, \Xi = \mathbb{R}^m$, and $f(x, \xi) = h(x) + B\xi$ where $h(x)$ is concave in x . Here $\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t = \gamma_1^\top \xi - \gamma_2 \|\xi\|_q$. Then

$$\psi_f(\gamma, \alpha, x) = \begin{cases} -\alpha h(x), & \text{if } \|\gamma_1 - \alpha B^\top\|_{\frac{q}{q-1}} \leq \gamma_2, \gamma_2 \geq 0 \\ +\infty, & \text{otherwise.} \end{cases}$$

Proposition 30. *Let*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^m) : \mathbb{E}_{\mathbb{P}}[\xi] = 0, \mathbb{E}_{\mathbb{P}}[\xi \xi^\top] \preceq \Sigma \right\},$$

with $\Xi = \mathbb{R}^m$, and $f(x, \xi) = (Ax + a)^\top \xi + Bx + b$. Here $\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t = \gamma_1^\top \xi - \langle \gamma_2, \xi \xi^\top \rangle$.

Then

$$\psi_f(\gamma, \alpha, x) = \begin{cases} -\alpha(Bx + b) + \min\{t : T(t, \gamma, \alpha, x) \succeq 0\} & \text{if } \gamma_2 \succeq 0 \\ +\infty & \text{otherwise,} \end{cases}$$

$$\text{with } T(t, \gamma, \alpha, x) = \begin{bmatrix} t & -\frac{1}{2}(\gamma_1 - \alpha(Ax + a))^\top \\ -\frac{1}{2}(\gamma_1 - \alpha(Ax + a)) & \gamma_2 \end{bmatrix}.$$

Proposition 31. *Let*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^m) : \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] = 0, \mathbb{E}_{\mathbb{P}}[(h^\top \boldsymbol{\xi})^2] \leq g(h^\top \mu)^2 \right\},$$

with $\Xi = \mathbb{R}^m$, and $f(x, \xi) = (Ax + a)^\top \xi + Bx + b$. Here $\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t = \gamma_1^\top \xi - \gamma_2 g(h^\top \xi)^2$.

Then

$$\psi_f(\gamma, \alpha, x) = \begin{cases} -\alpha(Bx + b) + \min \{t : T(t, \gamma, \alpha, x) \succeq 0\}, & \text{if } \gamma_2 \geq 0 \\ +\infty & \text{otherwise,} \end{cases}$$

$$\text{with } T(t, \gamma, \alpha, x) = \begin{bmatrix} t & -\frac{1}{2}(\gamma_1 - \alpha(Ax + a))^\top \\ -\frac{1}{2}(\gamma_1 - \alpha(Ax + a)) & \gamma_2 g h h^\top \end{bmatrix}.$$

The mean deviation ambiguity set in Proposition 29 has been studied in [47]. The first and second-moment ambiguity set in Proposition 30 has been studied in [30, 33, 51, 135], while the mean and coefficient of variation ambiguity set in Proposition 31 has been studied in [116].

5.3 Deterministic formulations

In this section we present two deterministic formulations associated with Z . The first is a direct reformulation using the strong duality result in Lemma 6. The second is an approximate formulation of Z via biconvex program which is shown to be nearly tight.

5.3.1 Direct reformulation

Lemma 6 is sufficient to derive the following deterministic reformulation of Z . We will investigate the convexity of this reformulation in Section 5.4.

Theorem 19. *Suppose Assumptions (A1) - (A4) hold, then*

$$Z = \left\{ x : \begin{aligned} & \lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq 1 - \epsilon, \\ & \lambda + \psi_0(\gamma) \leq 1, \\ & \lambda + \psi_{f_i}(\gamma, \alpha_i, x) \leq 0, \forall i \in I_1(x), \\ & \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_i \geq 0, \forall i \in I_1(x), \end{aligned} \right\} \quad \begin{aligned} (5.6a) \\ (5.6b) \\ (5.6c) \\ (5.6d) \end{aligned}$$

where the functions $\psi_0(\cdot), \psi(\cdot, \cdot, \cdot)$ are ϕ -conjugate functions as defined in Definition 8, and

$$I_1(x) := \{i \in [I] : \exists \xi \in \Xi, f_i(x, \xi) < 0\}.$$

Proof. Note that

$$Z_D = \left\{ x \in \mathbb{R}^n : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}[\mathbb{I}_+(F(x, \boldsymbol{\xi}))] \geq 1 - \epsilon \right\}.$$

By Lemma 6, the left-hand side of the constraint defining Z is equivalent to

$$\max \quad \lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \quad (5.7a)$$

$$\text{s.t.} \quad \lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t \leq \mathbb{I}_+(F(x, \xi)), \quad \forall \xi \in \Xi, \quad (5.7b)$$

$$\gamma_t \geq 0, t \in \mathcal{T}_2. \quad (5.7c)$$

From the definition of $\mathbb{I}_+(F(x, \xi))$, constraints (5.7b) are equivalent to

$$\lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t \leq 1, \quad \forall \xi \in \Xi, \quad (5.8a)$$

$$\lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t \leq 0, \quad \forall \xi \in \bigcup_{i \in [I]} \{\xi \in \Xi : f_i(x, \xi) < 0\}. \quad (5.8b)$$

Constraint (5.8a) is equivalent to

$$\lambda + \sup_{\xi \in \Xi} \left(\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t \right) \leq 1,$$

which is equivalent to (5.6b) by the definition of $\psi_0(\gamma)$. Next we focus on reformulating (5.8b).

Given x , let $\Xi_i(x) := \{\xi \in \Xi : f_i(x, \xi) < 0\}$ for all $i \in [I]$. Note that by definition of $I_1(x)$ we have that $\Xi_i(x) \neq \emptyset$ for all $i \in I_1(x)$. Thus (5.8b) is equivalent to

$$\lambda + \sup_{\xi \in \Xi_i(x)} \left(\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t \right) \leq 0, \quad \forall i \in I_1(x). \quad (5.9)$$

For any given x , since $f_i(x, \xi)$ and $\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t$ are continuous in ξ , Ξ is a closed set and by definition $\Xi_i(x) \neq \emptyset$ for each $i \in I_1(x)$, so we can replace the strict inequalities in $\Xi_i(x)$ by non-strict ones. Thus, (5.9) is equivalent to

$$\lambda + \sup_{\substack{\xi \in \Xi \\ f_i(x, \xi) \leq 0}} \left(\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t \right) \leq 0, \quad \forall i \in I_1(x). \quad (5.10)$$

For any given $\gamma \in \mathbb{R}^{|\mathcal{T}_1|} \times \mathbb{R}_+^{|\mathcal{T}_2|}$, $\lambda \in \mathbb{R}$ and $x \in S$, (5.10) implies that the following constraint system on ξ is insolvable

$$\lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t > 0, \xi \in \Xi, f_i(x, \xi) \leq 0, \quad (5.11)$$

for each $i \in I_1(x)$.

By definition of the set $I_1(x)$, we have that there exists $\bar{\xi} \in \Xi$ such that $f_i(x, \bar{\xi}) < 0$. Thus by Lemma 8, (5.11) is equivalent to that there exists an $\alpha_i \geq 0$,

$$\lambda + \sup_{\xi \in \Xi} \left(\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t - \alpha_i f_i(x, \xi) \right) \leq 0,$$

for each $i \in I_1(x)$. By definition of ψ_{f_i} the above system is equivalent to (5.6c). \square

We remark that reformulation (5.6) of Z is not convex since each function $\psi_{f_i}(\cdot)$ is not in general convex for $i \in [I]$, and also because the index set $I_1(x)$ depends on x . In the subsequent sections, we will explore the tractability of the set Z by establishing conditions under which $\{\psi_{f_i}(\gamma, \alpha_i, x)\}_{i \in [I]}$ are convex and $I_1(x)$ can be replaced by $[I]$.

We also remark that by Lemma 3.1. in [95], for any distribution $\mathbb{P} \in \mathcal{P}$, there exists a discrete distribution $\hat{\mathbb{P}} \in \mathcal{P}$ with a finite support of at most $1 + |\mathcal{T}_1| + |\mathcal{T}_2|$ points. Therefore, there exists a worst-case distribution \mathbb{P}^* which achieves the infimum in (5.1c) has a finite support with points $\{\xi_j^*\}_{j \in [J]}$ with $J \leq 1 + |\mathcal{T}_1| + |\mathcal{T}_2|$, i.e., \mathbb{P}^* satisfies

$$\mathbb{P}^*(\xi) = \sum_{j \in [J]} p_j^* \mathbb{I}(\xi = \xi_j^*).$$

We first observe that $\lambda + \psi_0(\gamma)$ can be lower bounded by $1 - \epsilon$.

Corollary 5. *For any $(\lambda, \gamma, \alpha, x)$ satisfying (5.6), we must have $\lambda + \psi_0(\gamma) \geq 1 - \epsilon$.*

Proof. For any given $\mathbb{P} \in \mathcal{P}$, (5.3) yields

$$\int_{\Xi} \phi_t(\xi) \mathbb{P}(d\xi) = g_t, \forall t \in \mathcal{T}_1, \int_{\Xi} \phi_t(\xi) \mathbb{P}(d\xi) \geq g_t, \forall t \in \mathcal{T}_2.$$

Since $\gamma_t \geq 0$ for each $t \in \mathcal{T}_2$ by aggregating the above inequalities with multipliers $\{\gamma_t\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$, we obtain

$$\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \int_{\Xi} \phi_t(\xi) \gamma_t \mathbb{P}(d\xi) \geq \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t. \quad (5.12)$$

Since

$$\psi_0(\gamma) := \sup_{\xi \in \Xi} \left(\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t \right) \geq \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \int_{\Xi} \phi_t(\xi) \gamma_t \mathbb{P}(d\xi),$$

therefore (5.12) and $\lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq 1 - \epsilon$ imply

$$\lambda + \psi_0(\gamma) \geq 1 - \epsilon.$$

□

We also observe that for each $i \in I_1(x)$, α_i must be strictly positive. This observation is key to the proofs of several main results in subsequent sections, for instance, it allows us to prove the convexity of the set Z when $\Xi = \mathbb{R}^m$ and $\{\phi_t(\xi)\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$ are all linear functions.

Corollary 6. *For any x satisfying (5.6), we must have $\alpha_i > 0$ for all $i \in I_1(x)$.*

Proof. If $I_1(x) = \emptyset$, then we are done. Now let us assume that $I_1(x) \neq \emptyset$. Suppose $\alpha_{i_0} = 0$ for some $i_0 \in I_1(x)$. Then from (5.6c), we have $\lambda + \psi_{f_{i_0}}(\gamma, 0, x) \leq 0$, which is equivalent to $\lambda + \psi_0(\gamma) \leq 0$. This yields a contradiction to Corollary 5 that $\lambda + \psi_0(\gamma) \geq 1 - \epsilon$. □

5.3.2 Biconvex approximation

Recently, in [30], [51] and [135], the authors derived **CVaR** approximation of a joint DRCCP with linear uncertain constraints, which yields an almost exact feasible region of a DRCCP. The construction of such an approximation scheme is outlined below. First, for any given positive vector $\alpha \in \mathbb{R}_{++}^I$, we can convert the joint chance constraints into a single one as

$$\begin{aligned} Z &= \left\{ x \in \mathbb{R}^n : \alpha \in \mathbb{R}_{++}^I, \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\xi : \max_{i \in [I]} \{\alpha_i (-f_i(x, \xi))\} \leq 0] \geq 1 - \epsilon \right\} \\ &= \left\{ x \in \mathbb{R}^n : \alpha \in \mathbb{R}_{++}^I, \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\xi : \max_{i \in [I]} \{\alpha_i (-f_i(x, \xi))\} > 0] \leq \epsilon \right\} \end{aligned}$$

where the second equality is due to $\mathbb{P}[\xi : \max_{i \in [I]} \{\alpha_i (-f_i(x, \xi))\} > 0] + \mathbb{P}[\xi : \max_{i \in [I]} \{\alpha_i (-f_i(x, \xi))\} \leq 0] = 1$. For any given probability measure $\mathbb{P} \in \mathcal{P}$, apply the **CVaR** approximation of [73] to the above chance constraint, which yields a conservative approximation of Z as

$$Z \supseteq \left\{ x : \alpha \in \mathbb{R}_{++}^I, \sup_{\mathbb{P} \in \mathcal{P}} \left\{ \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}} \left[\left(\max_{i \in [I]} \{\alpha_i (-f_i(x, \xi))\} - \beta \right)_+ \right] \right\} \right\} \leq 0 \right\}.$$

We further note that if we interchange the infimum with the supremum, then by standard minimax argument, set Z is further approximated by

$$\begin{aligned} Z &\supseteq \left\{ x : \alpha \in \mathbb{R}_{++}^I, \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{\epsilon} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left(\max_{i \in [I]} \{ \alpha_i (-f_i(x, \xi)) \} - \beta \right)_+ \right] \right\} \leq 0 \right\} \\ &\supseteq Z_C = \left\{ x : \alpha \in \mathbb{R}_{++}^I, \beta \in \mathbb{R}, \beta + \frac{1}{\epsilon} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left(\max_{i \in [I]} \{ \alpha_i (-f_i(x, \xi)) \} - \beta \right)_+ \right] \leq 0 \right\} \end{aligned} \quad (5.13)$$

where the second inclusion is because infimum might not be achieved by any β .

The relation (5.13) leads us to reformulate Z_C as a disjunction of two sets by distinguishing whether $\beta = 0$ or not.

Theorem 20. $Z_C = X_C \cup Y_C$, where

$$X_C = \{x : F(x, \xi) \geq 0, \forall \xi \in \Xi\} \quad (5.14)$$

and

$$Y_C = \left\{ x : \begin{aligned} &\lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq 1 - \epsilon, \\ &\lambda + \psi_0(\gamma) \leq 1, \\ &\lambda + \psi_{f_i}(\gamma, \alpha_i, x) \leq 0, \forall i \in [I], \\ &\gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_i \geq 0, \forall i \in [I]. \end{aligned} \right\} \quad (5.15a)$$

$$\quad (5.15b)$$

$$\quad (5.15c)$$

$$\quad (5.15d)$$

Proof. We separate the proof into three parts.

- (i) Note that in (5.13), we must have $\beta \leq 0$; otherwise, as $(\max_{i \in [I]} \{ \alpha_i (-f_i(x, \xi)) \} - \beta)_+ \geq 0$ for all $\xi \in \Xi$, thus the expectation in (5.13) is always nonnegative, which implies that the left-hand side of (5.13) is strictly positive, a contradiction.

- (ii) For any $x \in Z_C$, there exists $(\alpha, \beta) \in \mathbb{R}_{++}^I \times \mathbb{R}_-$ such that

$$\beta \epsilon + \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left(\max_{i \in [I]} \{ \alpha_i (-f_i(x, \xi)) \} - \beta \right)_+ \right] \leq 0.$$

Now we distinguish whether $\beta = 0$ or $\beta < 0$.

- (a) If $\beta = 0$, then (5.13) yields

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left(\max_{i \in [I]} \{ \alpha_i (-f_i(x, \xi)) \} \right)_+ \right] \leq 0 \quad (5.16)$$

which is equivalent to

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[F(x, \xi) \geq 0] = 1 \geq 1 - \epsilon.$$

By continuity of $F(x, \xi)$, $\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[F(x, \xi) \geq 0] = 1$ implies that $F(x, \xi) \geq 0$ for all $\xi \in \Xi$, which implies that the feasible solution x must be in X_C .

(b) Suppose $\beta < 0$. Divide (5.13) by $-\beta$ and add ϵ on both sides, then we have

$$\frac{1}{-\beta} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left(\max_{i \in [I]} \{\alpha_i (-f_i(x, \xi))\} - \beta \right)_+ \right] \leq \epsilon. \quad (5.17)$$

Since $\beta < 0$, we can redefine α_i as $\alpha_i/(-\beta)$ for each $i \in [I]$. Thus, (5.17) yields

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\left(\max_{i \in [I]} \{\alpha_i (-f_i(x, \xi))\} + 1 \right)_+ \right] \leq \epsilon.$$

Subtracting one on both sides and flipping the sign of inequality yields

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[1 - \left(\max_{i \in [I]} \{\alpha_i (-f_i(x, \xi))\} + 1 \right)_+ \right] \geq 1 - \epsilon. \quad (5.18)$$

By Lemma 6, for any given $\alpha \in \mathbb{R}_{++}^I$, the infimum in the left-hand side of (5.18) is equivalent to

$$\max \lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t, \quad (5.19a)$$

$$\text{s.t. } \lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t \leq 1 - \left(\max_{i \in [I]} \{\alpha_i (-f_i(x, \xi))\} + 1 \right)_+, \forall \xi \in \Xi, \quad (5.19b)$$

$$\gamma_t \geq 0, t \in \mathcal{T}_2. \quad (5.19c)$$

Since the maximum value of (5.19) should be no smaller than $1 - \epsilon$, therefore, we can replace the maximization over λ, γ by existing λ, γ . In addition, breaking down the maximum in (5.19b), (5.18) is equivalent to

$$\lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq 1 - \epsilon, \quad (5.20a)$$

$$\lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t \leq 1, \forall \xi \in \Xi, \quad (5.20b)$$

$$\lambda + \left[\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t - \alpha_i f_i(x, \xi) \right] \leq 0, \forall \xi \in \Xi, i \in [I], \quad (5.20c)$$

$$\gamma_t \geq 0, t \in \mathcal{T}_2, \quad (5.20d)$$

for some $\alpha \in \mathbb{R}_{++}^I$. By definition of ψ_f , (5.20a)-(5.20c) are equivalent to (5.15a)-(5.15c). Therefore, $(\lambda, \gamma, \alpha, x)$ satisfies (5.15). Thus, $x \in Y_C$.

This implies that $Z_C \subseteq X_C \cup Y_C$.

(iii) Now let $x \in X_C \cup Y_C$. If $x \in X_C$, then choose $\beta = 0, \alpha = \mathbf{e}$, thus $x \in Z_C$. If $x \in Y_C$, there exists $(\lambda', \gamma', \alpha')$ such that $(\lambda', \gamma', \alpha', x)$ satisfies (5.15) and we must have $\alpha' > 0$ from Corollary 6. In (5.13), let $\beta = 1, \alpha = \alpha'$. Then by Lemma 6, the dual reformulation (5.13) is equivalent to the set Z_C . Thus, $x \in Z_C$.

□

Remark 1. To solve (5.1) over set $S \cap Z_C$, one can optimize $c^\top x$ over $S \cap X_C$ and $S \cap Y_C$ separately, then choose the minimum value.

Remark 2. We note that the left-hand sides of the constraint system (5.15) are biconvex in α and (λ, γ, x) , i.e., for any given $\alpha \in \mathbb{R}_+^I$, they are convex in (λ, γ, x) , and also convex in α for any given (λ, γ, x) .

We observe that Y_C is quite similar to (5.6) except that index set $I_1(x)$ is equal to $[I]$ in (5.15). Indeed, we are able to show that Z is characterized by Y_C below.

Theorem 21. Let

$$\mathcal{C} := \left\{ x \in Z : \exists i \in [I], \inf_{\xi \in \Xi} f_i(x, \xi) = 0 \right\}.$$

If $\mathcal{C} = \emptyset$, then

$$Y_C = Z_C = Z.$$

Proof. Since Theorem 20 implies that $Y_C \subseteq Z_C \subseteq Z$, we only need to show that $Z \subseteq Y_C$.

We first rewrite set Y_C as

$$Y_C = \left\{ x : \begin{aligned} & \lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq 1 - \epsilon, & (5.21a) \\ & \lambda + \psi_0(\gamma) \leq 1, & (5.21b) \\ & \lambda + \psi_{f_i}(\gamma, \alpha_i, x) \leq 0, \forall i \in I_1(x), & (5.21c) \\ & \lambda + \psi_{f_i}(\gamma, \alpha_i, x) \leq 0, \forall i \in [I] \setminus I_1(x), & (5.21d) \\ & \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_i \geq 0, \forall i \in [I]. & (5.21e) \end{aligned} \right\}$$

By definition we have $[I] \setminus I_1(x) = \{i \in [I] : f_i(x, \xi) \geq 0, \forall \xi \in \Xi\}$. For each $i \in [I] \setminus I_1(x)$, (5.21d) is equivalent to

$$\lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t \leq \alpha_i f_i(x, \xi), \forall \xi \in \Xi.$$

In the above reformulation, by taking supremum over the left-hand side and using the fact that $\lambda + \psi_0(\gamma) \leq 1$, we observe that

$$Y_C \supseteq \left\{ x : \begin{aligned} & 1 \leq \alpha_i \inf_{\xi \in \Xi} f_i(x, \xi), \forall i \in [I] \setminus I_1(x), & (5.22) \\ & (5.21a), (5.21b), (5.21c), \\ & \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_i \geq 0, \forall i \in [I]. \end{aligned} \right\}$$

Using the fact that $\alpha_i > 0$ from the proof of Theorem 20, the right-hand side in (5.22) is equivalent to

$$Y_C \supseteq \hat{Y}_C := \left\{ x : \begin{aligned} & 0 < \inf_{\xi \in \Xi} f_i(x, \xi), \forall i \in [I] \setminus I_1(x), & (5.23) \\ & (5.21a), (5.21b), (5.21c), \\ & \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_i \geq 0, \forall i \in I_1(x). \end{aligned} \right\}$$

From definition of \hat{Y}_C , we have

$$Z \setminus Y_C \subseteq Z \setminus \hat{Y}_C \subseteq \mathcal{C} := \{x \in Z : \exists i \in [I], \inf_{\xi \in \Xi} f_i(x, \xi) = 0\}$$

where the first inclusion is due to $\hat{Y}_C \subseteq Y_C$, and the second inclusion is because for any $x \in Z$ but $x \notin \mathcal{C}$, we must have $x \in \hat{Y}_C$. \square

A direct observation from the proof of Theorem 21 is the sufficient conditions when α could be bounded. This observation is useful for binary DRCCP which will be discussed in subsequent sections.

Corollary 7. *If S is compact and $I_1(x) = [I]$ for all $x \in Z$ (i.e., $Z = Y_C$), then there exists an $M \in \mathbb{R}_{++}^I$ such that in $S \cap Y_C$, $\alpha_i \leq M_i$ for each $i \in [I]$.*

Proof. By Theorem 21, we have

$$Z \setminus Z_C \subseteq Z \setminus Y_C \subseteq \mathcal{C} := \{x \in Z : \exists i \in [I], \inf_{\xi \in \Xi} f_i(x, \xi) = 0\}.$$

If $I = 1$, then we have

$$\mathcal{C} \subseteq \{x \in \mathbb{R}^n : f_1(x, \xi) \geq 0, \forall \xi \in \Xi\} \subseteq X_C.$$

Thus,

$$Z \subseteq Z_C \cup X_C = Z_C,$$

i.e. $Z = Z_C$. □

Next we observe that $Z_C = Z$ when there is a single uncertain constraint.

Corollary 8. *When $I = 1$, we have $Z_C = Z$.*

Proof. By Theorem 21, we have

$$Y_C \subseteq Z_C \subseteq Z = \text{cl}(Y_C).$$

Thus,

$$Z \setminus Z_C \subseteq Z \setminus Y_C \subseteq \mathcal{C} := \{x \in \mathbb{R}^n : \exists i \in [I] \setminus I_1(x), \inf_{\xi \in \Xi} f_i(x, \xi) = 0\}$$

where the first inclusion is due to $Y_C \subseteq Z_C$, and the second inclusion is because for any $x \in Z$ but $x \notin \mathcal{C}$, we must have $x \in Y_C$.

If $I = 1$, then we have

$$\mathcal{C} \subseteq \{x \in \mathbb{R}^n : f_1(x, \xi) \geq 0, \forall \xi \in \Xi\} \subseteq X_C.$$

Thus,

$$Z \subseteq Z_C \cup X_C = Z_C,$$

i.e. $Z = Z_C$. □

Remark 3. *A special case of Corollary 8 has been observed by [123, 135] for a DRCCP with first- and second- moment constraints. Here, we provide a different proof and our results apply to a DRCCP with more general convex moment constraints.*

Despite the tightness of Z_C , due to the biconvex terms in set Y_C , it is nonconvex in general. However, as shown in Sections 5.4 and 5.5, in some cases, it is possible that these biconvex terms can be convexified.

5.4 Convexity conditions for Z

In this section, we will explore some settings under which the set Z is convex. The results in the first two subsections are derived by constructing a new formulation which projects out dual variables λ, α in (5.6) and proving that the new formulation is convex and equivalent to the set Z . The remaining two results are to explore the positive homogeneity of mappings $\{\phi_t(\xi)\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$ or $F(x, \xi)$.

5.4.1 Single uncertain constraint

We show that if there is a single uncertain constraint (i.e., $I = 1$), then the set Z is convex.

Theorem 22. *When $I = 1$, then*

$$Z = Z_C = \left\{ x : \begin{array}{l} - \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t + (1 - \epsilon) \psi_0(\gamma) + \epsilon \psi_{f_1}(\gamma, 1, x) \leq 0, \\ \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \end{array} \right\} \quad (5.24a)$$

$$(5.24b)$$

which is a convex set.

Proof. Note that $Z_C = Z$ from Corollary 8. By definition of the set Z_C in (5.13), we can obtain

$$Z_C = \left\{ x : \alpha_1 > 0, \beta \in \mathbb{R}, \beta + \frac{1}{\epsilon} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [(\alpha_1 (-f_1(x, \xi)) - \beta)_+] \leq 0 \right\} \quad (5.25)$$

By replacing β with $-\beta$ and flipping the sign of inequality, we have that (5.25) is equivalent to

$$Z_C = \left\{ x : \alpha_1 > 0, \beta \in \mathbb{R}, \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [\min(\alpha_1 f_1(x, \boldsymbol{\xi}), \beta)] \geq (1 - \epsilon)\beta \right\} \quad (5.26)$$

Then by Lemma 6 and the proof of Theorem 20, the infimum in the left-hand side of (5.26) is equivalent to

$$Z_C = \left\{ x : \begin{aligned} & \lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq (1 - \epsilon)\beta \\ & \lambda + \psi_0(\gamma) \leq \beta \\ & \lambda + \psi_{f_1}(\gamma, \alpha_1, x) \leq 0, \\ & \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_1 > 0. \end{aligned} \right\} \quad (5.27a)$$

$$(5.27b)$$

$$(5.27c)$$

$$(5.27d)$$

Next, we project out variables λ, β from set Z_C in (5.27) by Fourier-Motzkin elimination procedure and obtain

$$Z_C = \left\{ x : \begin{aligned} & - \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t + (1 - \epsilon)\psi_0(\gamma) + \epsilon\psi_{f_1}(\gamma, \alpha_1, x) \leq 0, \\ & \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_1 > 0. \end{aligned} \right\} \quad (5.28a)$$

$$(5.28b)$$

We note that in (5.28a), α_1 must be positive and finite. Thus, by scaling it to be 1, we obtain (5.24).

Since $\phi_0(\cdot)$ and $\phi_{f_1}(\cdot, 1, \cdot)$ are convex functions, Z is a convex set. \square

Theorem 22 implies that for a single DRCCP, Z can be always reformulated as a convex set. This has been observed by [51, 123, 135] where \mathcal{P} is only constrained by first- and second-moments. Here, we extend this result to more general moment constraints.

5.4.2 Single moment constraint

Here we consider the case that \mathcal{P} is described by a single moment constraint. First, we show that for one inequality constraint (i.e., $|\mathcal{T}_1| = 0, |\mathcal{T}_2| = 1$), the set Z is convex. The main idea behind the proof is to project out the α variables from characterization (5.6) in Theorem 19.

Theorem 23. If $|\mathcal{T}_1| = 0$ and $|\mathcal{T}_2| = 1$, then

$$Z = \left\{ x : \begin{array}{l} -g_1 \hat{\gamma}_i + (1 - \epsilon) \psi_0(\hat{\gamma}_i) + \epsilon \psi_{f_i}(\hat{\gamma}_i, 1, x) \leq 0, \forall i \in [I], \\ \hat{\gamma}_i \geq 0, \forall i \in [I], \end{array} \right\} \quad (5.29a)$$

$$(5.29b)$$

which is a convex set.

Proof. Let Z^* be the set defined on the right-hand side of (5.29), which is clearly convex. We will first show that set Z is equivalent to \tilde{Z} by projecting out dual multiplier λ and aggregating two types of constraints into one. Next, we show that the consolidated set \tilde{Z} is equivalent to the convex set Z^* . The proof proceeds in three steps.

(a) First, using Fourier-Motzkin method to project out variable λ in (5.6), we can reformulate Z as

$$Z = \left\{ \begin{array}{l} -g_1 \gamma_1 + \psi_0(\gamma_1) \leq \epsilon, \\ x : -g_1 \gamma_1 + \psi_{f_i}(\gamma_1, \alpha_i, x) \leq \epsilon - 1, \forall i \in I_1(x), \\ \gamma_1 \geq 0, \alpha_i > 0, \forall i \in [I], \end{array} \right\} \quad (5.30a)$$

$$(5.30b)$$

$$(5.30c)$$

where in (5.30c), we let $\alpha_i > 0$ for all $i \in [I]$ due to Corollary 6.

Next, we can get a relaxation of the Z by aggregating the two constraints (5.30a), (5.30b) above together as $(1 - \epsilon) \times (5.30a) + \epsilon \times (5.30b)$:

$$\tilde{Z} = \left\{ x : \begin{array}{l} -g_1 \gamma_1 + (1 - \epsilon) \psi_0(\gamma_1) + \epsilon \psi_{f_i}(\gamma_1, \alpha_i, x) \leq 0, \forall i \in I_1(x), \\ \gamma_1 \geq 0, \alpha_i > 0, \forall i \in [I]. \end{array} \right\} \quad (5.31a)$$

$$(5.31b)$$

Set \tilde{Z} is a relaxation of Z , i.e. $Z \subseteq \tilde{Z}$.

Next we show that $\tilde{Z} \subseteq Z$. Recall that $I_1(x) := \{i \in [I] : \exists \xi \in \Xi, f_i(x, \xi) < 0\}$. Given a point $x \in \tilde{Z}$, we consider two cases $I_1(x) = \emptyset$ and $I_1(x) \neq \emptyset$. If $I_1(x) = \emptyset$, then let $\gamma_1 = 0$ and $\alpha = \mathbf{e}$ (\mathbf{e} is the all-one vector). Clearly (γ_1, α, x) satisfies constraints in (5.30). Hence $x \in Z$.

Now suppose that $I_1(x) \neq \emptyset$. As $x \in \tilde{Z}$, there exists (γ_1, α) such that (γ_1, α, x) satisfies constraints in (5.31). First of all, we claim that $\gamma_1 > 0$; otherwise, suppose that $\gamma_1 = 0$,

then by (5.31a), Definition 8 and the fact that $\alpha \in \mathbb{R}_{++}^I$, we have

$$\sup_{\xi \in \Xi} -\alpha_i f_i(x, \xi) \leq 0,$$

which implies that $f_i(x, \xi) \geq 0$ for all $\xi \in \Xi$ and $i \in I_1(x)$, contradicting $I_1(x) \neq \emptyset$. Thus, we must have $\gamma_1 > 0$.

Next, we claim that $-g_1\gamma_1 + \psi_0(\gamma_1) > 0$. Indeed, by Assumption (A4), there exists a probability measure $\mathbb{P} \in \mathcal{P}_0(\Xi)$ such that $\mathbb{E}_{\mathbb{P}}[\phi_1(\xi)] > g_1$. Thus

$$g_1 < \mathbb{E}_{\mathbb{P}}[\phi_1(\xi)] \leq \sup_{\xi \in \Xi} \phi_1(\xi).$$

Since $\gamma_1 > 0$, we must have

$$-g_1\gamma_1 + \psi_0(\gamma_1) = \gamma_1[-g_1 + \sup_{\xi \in \Xi} \phi_1(\xi)] > 0.$$

Define $\bar{\gamma}_1 = \epsilon\gamma_1 / [-g_1\gamma_1 + \psi_0(\gamma_1)]$. Thus,

$$\begin{aligned} -g_1\bar{\gamma}_1 + \psi_0(\bar{\gamma}_1) &= -g_1\epsilon\gamma_1 / [-g_1\gamma_1 + \psi_0(\gamma_1)] + \sup_{\xi \in \Xi} (\epsilon\gamma_1\phi_1(\xi) / [-g_1\gamma_1 + \psi_0(\gamma_1)]) \\ &= \frac{\epsilon}{-g_1\gamma_1 + \psi_0(\gamma_1)} [-g_1\gamma_1 + \psi_0(\gamma_1)] = \epsilon, \end{aligned}$$

where the first and second equalities are from the definition of $\psi_0(\cdot)$ and construction of $\bar{\gamma}_1$; and similarly for each $i \in I_1(x)$,

$$\begin{aligned} -g_1\bar{\gamma}_1 + \psi_{f_1}(\bar{\gamma}_1, \alpha_i, x) &= -g_1\epsilon\gamma_1 / [-g_1\gamma_1 + \psi_0(\gamma_1)] \\ &\quad + \sup_{\xi \in \Xi} (\epsilon\gamma_1\phi_1(\xi) / [-g_1\gamma_1 + \psi_0(\gamma_1)] - \epsilon\alpha_i f_i(x, \xi) / [-g_1\gamma_1 + \psi_0(\gamma_1)]) \\ &= \frac{\epsilon}{-g_1\gamma_1 + \psi_0(\gamma_1)} [-g_1\gamma_1 + \psi_{f_1}(\gamma, \alpha_i, x)] \\ &\leq \frac{\epsilon}{-g_1\gamma_1 + \psi_0(\gamma_1)} \left[-g_1\gamma_1 + \frac{g_1\gamma_1 - (1-\epsilon)\psi_0(\gamma)}{\epsilon} \right] = \epsilon - 1, \end{aligned}$$

where the first and second equalities are from the definition of $\psi_0(\cdot)$ and construction of $\bar{\gamma}_1$, and the last inequality is due to (5.31a).

Hence, $(\bar{\gamma}_1, \alpha, x)$ satisfy the constraints in (5.30); i.e., $x \in Z$. Thus, $\tilde{Z} = Z$.

(b) Now we show that $Z^* \subseteq \tilde{Z}$. Given a point $x \in Z^*$, if $I_1(x) = \emptyset$, then clearly $x \in \tilde{Z}$.

Suppose $I_1(x) \neq \emptyset$. Since $x \in Z^*$, there must exist a vector $\hat{\gamma}$ such that $(\hat{\gamma}, x)$ satisfies (5.29). For any $i \in I_1(x)$, we must have $\hat{\gamma}_i > 0$; otherwise, (5.31a) yields $\psi_{f_i}(0, 1, x) \leq 0$; i.e., $f_i(x, \xi) \geq 0$ for all $\xi \in \Xi$, contradicting $i \in I_1(x)$.

Let $\gamma_1 = \max_{i \in I_1(x)} \hat{\gamma}_i$ and set $\alpha_i = \gamma_1 / \hat{\gamma}_i$ for each $i \in I_1(x)$ and $\alpha_i = 1$, otherwise. Then for each $i \in I_1(x)$,

$$\begin{aligned} & -g_1\gamma_1 + (1-\epsilon)\psi_0(\gamma_1) + \epsilon\psi_{f_i}(\gamma_1, \alpha_i, x) \\ &= -g_1\gamma_1 + (1-\epsilon)\sup_{\xi \in \Xi}(\gamma_1\phi_1(\xi)) + \epsilon\sup_{\xi \in \Xi}(\gamma_1\phi_1(\xi) - \gamma_1/\hat{\gamma}_i f_i(x, \xi)) \\ &= \frac{\gamma_1}{\hat{\gamma}_i} \left[-g_1\hat{\gamma}_i + (1-\epsilon)\sup_{\xi \in \Xi}(\hat{\gamma}_i\phi_1(\xi)) + \epsilon\sup_{\xi \in \Xi}(\hat{\gamma}_i\phi_1(\xi) - f_i(x, \xi)) \right] \\ &= \frac{\gamma_1}{\hat{\gamma}_i} [-g_1\hat{\gamma}_i + (1-\epsilon)\psi_0(\hat{\gamma}_i) + \epsilon\psi_{f_i}(\hat{\gamma}_i, 1, x)] \leq 0, \end{aligned}$$

where the first three equalities are from the definition of $\psi(\cdot)$ and construction of γ_1, α , and the last inequality due to (5.29a) and the fact that $\gamma_1 = \max_{j \in I_1(x)} \hat{\gamma}_j \geq \hat{\gamma}_i > 0$.

Thus, (γ_1, α, x) satisfies constraints (5.31), i.e. $x \in \tilde{Z}$.

- (c) Next we show that $\tilde{Z} \subseteq Z^*$. Given $x \in \tilde{Z}$, there exists (γ_1, α) such that (γ_1, α, x) satisfies constraints in (5.31). For each $i \in [I] \setminus I_1(x)$, let $\hat{\gamma}_i = 0$; otherwise, let $\hat{\gamma}_i = \gamma_1 / \alpha_i$.

Then for each $i \in [I] \setminus I_1(x)$, we have

$$-g_1\hat{\gamma}_i + (1-\epsilon)\psi_0(\hat{\gamma}_i) + \epsilon\psi_{f_i}(\hat{\gamma}_i, 1, x) = \epsilon\sup_{\xi \in \Xi}(-f_i(x, \xi)) \leq 0,$$

where the first equality is from the definition of $\psi(\cdot, \cdot, \cdot)$ and $\hat{\gamma}_i = 0$, and the first inequality is due to $i \in [I] \setminus I_1(x)$, thus $f_i(x, \xi) \geq 0$ for all $\xi \in \Xi$. On the other hand, for for each $i \in I_1(x)$, we have

$$\begin{aligned} & -g_1\hat{\gamma}_i + (1-\epsilon)\psi_0(\hat{\gamma}_i) + \epsilon\psi_{f_i}(\hat{\gamma}_i, 1, x) \\ &= -g_1\gamma_1/\alpha_i + (1-\epsilon)\sup_{\xi \in \Xi}(\phi_1(\xi)\gamma_1/\alpha_i) + \epsilon\sup_{\xi \in \Xi}(\phi_1(\xi)\gamma_1/\alpha_i - f_i(x, \xi)) \\ &= \frac{1}{\alpha_i} \left[-g_1\gamma_1 + (1-\epsilon)\sup_{\xi \in \Xi}(\gamma_1\phi_1(\xi)) + \epsilon\sup_{\xi \in \Xi}(\gamma_1\phi_1(\xi) - \alpha_i f_i(x, \xi)) \right] \leq 0, \end{aligned}$$

where the first two equalities are due to the definition of $\psi(\cdot, \cdot, \cdot)$ and $\hat{\gamma}_i = \gamma_1 / \alpha_i$, and the last inequality is due to (5.31a) and $\alpha_i > 0$.

Thus, $(\hat{\gamma}, x)$ satisfies the constraints in (5.29); i.e., $x \in \tilde{Z}$.

Thus, $Z = \tilde{Z} = Z^*$. Since $\phi_0(\cdot)$ and $\{\phi_{f_i}(\cdot, 1, \cdot)\}_{i \in [I]}$ are convex functions, Z is a convex set. \square

We remark that the proof of Theorem 23 only holds for the case of one moment inequality. If there is more than one moment inequality, it is difficult to project out the dual multipliers $\{\alpha_i\}_{i \in [I]}$.

Another observation is that in the reformulation (5.29), the constraints (5.29a) are I replications of (5.24a). Indeed, let us consider the following set

$$Z_O = \left\{ x : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\xi : f_i(x, \xi) \geq 0] \geq 1 - \epsilon, \forall i \in [I] \right\} \quad (5.32)$$

which relaxes the requirement to satisfy all uncertain constraints, and is an outer approximation of Z . We can show that the relaxed set Z_O is equivalent to Z if there is only one moment inequality in \mathcal{P} .

Proposition 32. *If $|\mathcal{T}_1| = 0$ and $|\mathcal{T}_2| = 1$, then $Z_O = Z$.*

Proof. Since set Z_O consists of I single DRCCP, therefore Theorem 22, Z_O is equivalent to

$$Z_O = \left\{ x : \begin{array}{l} -g_1\gamma_{1i} + (1 - \epsilon)\psi_0(\gamma_{1i}) + \epsilon\psi_{f_i}(\gamma_{1i}, 1, x) \leq 0, \forall i \in [I], \\ \gamma_{1i} \geq 0, \forall i \in [I], \end{array} \right\}$$

which clearly equals to (5.29) defining set Z . \square

This result in Proposition 32 does not hold for general ambiguity set \mathcal{P} . Consider the following example:

Example 5. *Let $n = 1, I = 2$ and $f_1(x, \xi) = \xi x + T, f_2(x, \xi) = -\xi x + T$ and*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}) : \mathbb{E}_{\mathbb{P}}[\xi] = 0, \mathbb{E}_{\mathbb{P}}[\xi^2] \leq 1 \right\},$$

then according to Theorem 3.1 [24] and Theorem 2 [117], we can reformulate sets Z_O and Z as follows:

$$Z_O = \left\{ x : |x| \leq \sqrt{\frac{\epsilon}{1 - \epsilon}} T \right\},$$

$$Z = \{x : |x| \leq \sqrt{\epsilon}T\}.$$

Clearly, when $\epsilon \rightarrow 1$, $Z_O \rightarrow \mathbb{R}$ but Z is always bounded. Hence the distance between Z and Z_O can be arbitrarily large.

Note that if there is one linear moment equality, we can reformulate Z as a disjunction of two sets by treating an equality constraint as two inequalities, then applying the same technique of Theorem 23.

Theorem 24. *If $|\mathcal{T}_1| = 1$ and $|\mathcal{T}_2| = 0$, then*

$$Z = \bar{Z}^1 \cup \bar{Z}^2,$$

where

$$\bar{Z}^1 = \left\{ x : \begin{array}{l} -g_1\hat{\gamma}_i + (1-\epsilon)\psi_0(\hat{\gamma}_i) + \epsilon\psi_{f_i}(\hat{\gamma}_i, 1, x) \leq 0, \forall i \in [I], \\ \hat{\gamma}_i \geq 0, \forall i \in [I], \end{array} \right\} \quad (5.33a)$$

$$(5.33b)$$

$$\bar{Z}^2 = \left\{ x : \begin{array}{l} -g_1\hat{\gamma}_i + (1-\epsilon)\psi_0(\hat{\gamma}_i) + \epsilon\psi_{f_i}(\hat{\gamma}_i, 1, x) \leq 0, \forall i \in [I], \\ \hat{\gamma}_i \leq 0, \forall i \in [I], \end{array} \right\} \quad (5.34a)$$

$$(5.34b)$$

and \bar{Z}^1, \bar{Z}^2 are convex sets.

Proof. From the proof of Theorem 23, we know

$$Z = \left\{ x : \begin{array}{l} -g_1\gamma_1 + (1-\epsilon)\psi_0(\gamma_1) + \epsilon\psi_{f_i}(\gamma_1, \alpha_i, x) \leq 0, \forall i \in I_1(x), \\ \alpha_i \geq 0, \forall i \in [I], \end{array} \right\} \quad (5.35a)$$

$$(5.35b)$$

and if $\gamma_1 \geq 0$, then $Z = Z^1$, while if $\gamma_1 \leq 0$, then $Z = Z^2$. The conclusion follows by combining these two sets together with disjunction. \square

Different from a result in [47], where the authors only showed tractability of a linear DRCCP with right-hand-side uncertainty (i.e., $f_i(x, \xi) = a_i^\top x - \xi^\top b^i$ with constants $a_i \in \mathbb{R}^n, b^i \in \mathbb{R}^m$), our results in Theorems 23 and 24 apply to both right-hand-side and left-hand-side uncertainty.

The following example shows an application of Theorem 23.

Example 6. Consider a stochastic multi-dimensional continuous knapsack problem. There are n items and I knapsacks, where for each item j , c_j represents its value, and ξ_i represents i th knapsack's random weight vector, and let b_i be the total capacity of i th knapsack. The variable x_j denotes the portion of j th item being picked. Suppose that we know the total absolute deviation of weight, thus \mathcal{P} is defined as

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\|\tilde{\xi}\|_1] \leq g \right\},$$

where $\Xi = \mathbb{R}^{n \times I}$ and $\xi_i = \tilde{\xi}_i + \mu_i$ for all $i \in [I]$.

Now the distributionally robust multi-dimensional continuous knapsack problem is formulated as

$$\begin{aligned} v^* &= \max_{x \in [0,1]^n} c^\top x, \\ \text{s.t. } & \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[x^\top(\mu_i + \tilde{\xi}_i) \leq b_i, \forall i \in [I]] \geq 1 - \epsilon. \end{aligned} \quad (5.36)$$

Note that for each $i \in [I]$, the i th uncertain constraint is $f_i(x, \xi) = b_i - x^\top \mu_i - x^\top \tilde{\xi}_i$. By Proposition 27 with $q = 1$, we have

$$\psi_{f_i}(\gamma, \alpha_i, x) = \begin{cases} -\alpha_i(b_i - \mu_i^\top x), & \text{if } \|\alpha_i x\|_\infty \leq \gamma, \\ +\infty, & \text{otherwise,} \end{cases}$$

and $\psi_0(\gamma) = 0$.

Thus, according to Theorem 23, Problem (5.36) is equivalent to the following linear program

$$v^* = \max c^\top x, \quad (5.37a)$$

$$\text{s.t. } x \in [0, 1]^n, \quad (5.37b)$$

$$g\gamma_i \leq \epsilon(b_i - \mu_i^\top x), \forall i \in [I], \quad (5.37c)$$

$$\|x\|_\infty \leq \gamma_i, \forall i \in [I], \quad (5.37d)$$

$$\gamma_i \geq 0, \forall i \in [I]. \quad (5.37e)$$

5.4.3 Linear moment constraints

Here, we show that if $\Xi = \mathbb{R}^m$ and the ambiguity set is defined only by linear moment constraints, then the set Z is equal to X_C as defined in (5.14). Hence, DRCCP is equivalent

to a robust convex program.

Theorem 25. Suppose $\Xi = \mathbb{R}^m$ and $\{\phi_t(\xi)\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$ are all linear functions, then

$$Z = X_C = \{x \in \mathbb{R}^n : f_i(x, \xi) \geq 0, \forall \xi \in \mathbb{R}^m, i \in [I]\}. \quad (5.38)$$

Proof. Since $\{\phi_t(\xi)\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$ are all linear functions, let $\phi_t(\xi) = w_t^\top \xi$ for all $t \in \mathcal{T}_1 \cup \mathcal{T}_2$. We only need to show that $I_1(x) = \emptyset$ for any given $x \in Z$, where $I_1(x) := \{i \in [I] : \exists \xi \in \Xi, f_i(x, \xi) < 0\}$.

Suppose that $I_1(x) \neq \emptyset$ for some $x \in Z$. Since $\phi_t(\xi) = w_t^\top \xi$ for all $t \in \mathcal{T}_1 \cup \mathcal{T}_2$ and $\Xi = \mathbb{R}^m$, (5.6b) in Theorem 19 yields have

$$\lambda + \sup_{\xi \in \mathbb{R}^m} \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} w_t^\top \xi \gamma_t \leq 1,$$

which implies that

$$\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} w_t \gamma_t = 0.$$

Meanwhile, in (5.3), for any $\mathbb{P} \in \mathcal{P}_0(\Xi)$, we have $\mathbb{E}_{\mathbb{P}}[w_t^\top \xi] = g_t$ for all $t \in \mathcal{T}_1$ and $\mathbb{E}_{\mathbb{P}}[w_t^\top \xi] \geq g_t$ for all $t \in \mathcal{T}_2$. Multiplying these equalities and inequalities with $\{\gamma_t\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$ (note that $\gamma_t \geq 0$ for each $t \in \mathcal{T}_2$), we have

$$\mathbb{E}_{\mathbb{P}} \left[\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \gamma_t w_t^\top \xi \right] \geq \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t,$$

i.e., $\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \leq 0$ as $\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} w_t \gamma_t = 0$.

Now from (5.6c) in Theorem 19, for each $i \in I_1(x)$, we have

$$\lambda + \psi_{f_i}(\gamma, \alpha_i, x) = \lambda + \sup_{\xi \in \Xi} \left(\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \gamma_t w_t^\top \xi - \alpha_i f_i(x, \xi) \right) = \lambda + \sup_{\xi \in \Xi} (-\alpha_i f_i(x, \xi)),$$

where the first equality is due to the definition of $\phi(\cdot, \cdot, \cdot)$ and the second equality is because of $\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} w_t \gamma_t = 0$. As we know $i \in I_1(x)$ and $\alpha_i > 0$ from Corollary 6, we must have $\sup_{\xi \in \Xi} (-\alpha_i f_i(x, \xi)) > 0$. Hence, (5.6c) (i.e., $\lambda + \psi_{f_i}(\gamma, \alpha_i, x) \leq 0$) implies that $\lambda < 0$.

On the other hand, (5.6a) (i.e., $\lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq 1 - \epsilon$) implies that $\lambda \geq 1 - \epsilon$ since $\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \leq 0$. Thus, we have a contradiction. \square

This proposition suggests us that only considering first-moment information might not provide us a sufficient characterization of the ambiguous set and hence more nonlinear moment constraints are needed for a more realistic reformulation.

5.4.4 Nonlinear positively homogeneous moment constraints

Now we consider the case of multiple (possibly, nonlinear) moment constraints. Let us begin with the following technical lemma.

Lemma 10. *Suppose that Ξ is a closed convex cone, $\phi_t(\xi)$ is positively homogeneous on Ξ for each $t \in \mathcal{T}_2$, then Z is equivalent to*

$$Z = \left\{ x \in \mathbb{R}^n : \begin{aligned} & \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq -\epsilon, \\ & \bar{\psi}_0(\gamma) \leq 0, \\ & 1 + \psi_{f_i}(\gamma, \alpha_i, x) \leq 0, \forall i \in I_1(x), \\ & \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_i \geq 0, \forall i \in I_1(x), \end{aligned} \right\} \quad \begin{aligned} (5.39a) \\ (5.39b) \\ (5.39c) \\ (5.39d) \end{aligned}$$

where convex mapping $\bar{\psi}_0$ describes the domain of ϕ_0 , i.e.,

$$\psi_0(\gamma) = \begin{cases} 0, & \text{if } \bar{\psi}_0(\gamma) \leq 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

Proof. Since $\phi_t(\xi)$ is linear for each $t \in \mathcal{T}_1$ and is positively homogeneous on Ξ for each $t \in \mathcal{T}_2$, by part (iii) of Lemma 9, (5.6) reduces to

$$Z = \left\{ x : \begin{aligned} & \lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq 1 - \epsilon, \\ & \lambda \leq 1, \\ & \lambda + \psi_{f_i}(\gamma, \alpha_i, x) \leq 0, \forall i \in I_1(x), \\ & (5.39b), (5.39d). \end{aligned} \right\} \quad \begin{aligned} (5.40a) \\ (5.40b) \\ (5.40c) \end{aligned}$$

It remains to show that for any x in Z , we always have $\lambda = 1$.

Consider an $x \in Z$ with $(\lambda, \gamma, \alpha)$ such that $(\lambda, \gamma, \alpha, x)$ satisfies (5.40) and $\lambda < 1$. By Corollary 5 and $\psi_0(\gamma) = 0$ in its domain, we must have $\lambda \geq 1 - \epsilon$. Now construct a new

solution $(\bar{\lambda}, \bar{\gamma}, \bar{\alpha}, x)$ as $\bar{\lambda} = 1, \bar{\gamma} = \frac{\gamma}{\lambda}, \bar{\alpha} = \frac{\alpha}{\lambda}$. Clearly, $(\bar{\lambda}, \bar{\gamma}, \bar{\alpha}, x)$ also satisfies (5.40). Thus, we can always set $\lambda = 1$ in (5.40), which yields (5.39).

By Lemma 9, $\psi_0(\gamma)$ is convex in λ , therefore $\bar{\psi}_0$ is a convex mapping. \square

Next, we identify sufficient conditions for the set Z to be convex if the moment constraints are defined by positively homogeneous functions $\{\phi_t(\xi)\}_{t \in \mathcal{T}_2}$.

Theorem 26. *Suppose that Ξ is a closed convex cone, $\phi_t(\xi)$ is positively homogeneous on Ξ for each $t \in \mathcal{T}_2$, and for each $i \in [I]$, $f_i(x, \xi) = w_i(x) - h_i(\xi)$, where $w_i(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is a concave function and $h_i(\xi) : \Xi \rightarrow \mathbb{R}$ is a concave function and positively homogeneous on Ξ . Then Z is formulated as the following convex set*

$$Z = \left\{ x : \begin{array}{l} \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq -\epsilon, \\ 1 \leq \alpha_i w_i(x), \forall i \in [I] \setminus I_2, \\ 0 \leq w_i(x), \forall i \in I_2, \\ \bar{\psi}_0(\gamma) \leq 0, \\ \bar{\psi}_{-h_i}(\gamma, -\alpha_i) \leq 0, \forall i \in [I] \setminus I_2, \\ \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_i \geq 0, \forall i \in [I] \setminus I_2, \end{array} \right\} \quad \begin{array}{l} (5.41a) \\ (5.41b) \\ (5.41c) \\ (5.41d) \\ (5.41e) \\ (5.41f) \end{array}$$

where $I_2 := \{i \in [I] : h_i(\xi) \leq 0, \forall \xi \in \Xi\}$, and for each $i \in [I]$,

$$\psi_{f_i}(\gamma, \alpha_i, x) = -\alpha_i w_i(x) + \psi_{-h_i}(\gamma, \alpha_i, x) = \begin{cases} -\alpha_i w_i(x), & \text{if } \bar{\psi}_{-h_i}(\gamma, \alpha_i) \leq 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

with the convex mapping $\bar{\psi}_{-h_i}(\gamma, \alpha_i)$.

Proof. We first note that $0 \leq w_i(x)$ for all $i \in I_2$ for each $x \in Z$. We prove it by contradiction. Suppose that $I_1(x) \cap I_2 \neq \emptyset$. Thus, in (5.39c), for each $i \in I_1(x) \cap I_2$, we have

$$\begin{aligned} 0 &\geq 1 + \psi_{f_i}(\gamma, \alpha_i, x) = \psi_{f_i}(\gamma, \alpha_i, x) = 1 - \alpha_i w_i(x) + \psi_{-h_i}(\gamma, \alpha_i, x) \\ &= \begin{cases} 1 - \alpha_i w_i(x), & \text{if } \bar{\psi}_{-h_i}(\gamma, \alpha_i) \leq 0, \\ +\infty, & \text{otherwise,} \end{cases} \end{aligned}$$

where the last inequality due to positive homogeneity of $\{\phi_t(\xi)\}_{t \in \mathcal{T}_1 \cup \mathcal{T}_2}$ and $h_i(\xi)$. Thus, we must have $\bar{\psi}_{-h_i}(\gamma, \alpha_i) \leq 0$ and $\alpha_i w_i(x) \geq 1$. This implies that $w_i(x) \geq \sup_{\xi} h_i(\xi)$, i.e. $f_i(x, \xi) = w_i(x) - h_i(\xi) \geq 0$ for all $\xi \in \Xi$. Therefore, $i \notin I_1(x)$, contradiction.

Thus, set Z is now equivalent to $\{0 \leq w_i(x), \forall i \in I_2\} \cap \tilde{Z}$, where

$$\tilde{Z} = \left\{ x \in \mathbb{R}^n : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[\xi : w_i(x) \geq h_i(\xi), \forall i \in [I] \setminus I_2] \geq 1 - \epsilon \right\}$$

Now from the proof in Theorem 20, we know

$$\tilde{Z} \setminus \tilde{Y}_C \subseteq \mathcal{C} := \{x \in \tilde{Z} : \exists i \in [I] \setminus I_2, w_i(x) = \sup_{\xi \in \Xi} h_i(\xi)\},$$

where

$$\tilde{Y}_C = \left\{ x \in \mathbb{R}^n : \begin{aligned} & \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq -\epsilon, \\ & \lambda + \psi_0(\gamma) \leq 1, \\ & \lambda + \psi_{f_i}(\gamma, \alpha_i, x) \leq 0, \forall i \in [I] \setminus I_2, \\ & \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_i \geq 0, \forall i \in [I] \setminus I_2. \end{aligned} \right\} \quad (5.42)$$

Since for each $i \in [I] \setminus I_2$, we have $h_i(\xi) > 0$ for some $\xi \in \Xi$, thus by positive homogeneity, we must have $\sup_{\xi \in \Xi} h_i(\xi) = \infty$. Hence, $\mathcal{C} = \emptyset$ and $\tilde{Y}_C = \tilde{Z}$. By Lemma 10, we can $\lambda = 1$ in set \tilde{Y}_C . Thus,

$$Z = \{0 \leq w_i(x), \forall i \in I_2\} \cap \tilde{Z} = \{0 \leq w_i(x), \forall i \in I_2\} \cap \tilde{Y}_C$$

leads to

$$Z = \left\{ x \in \mathbb{R}^n : \begin{aligned} & \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq -\epsilon, \\ & \bar{\psi}_0(\gamma) \leq 0, \\ & 0 \leq w_i(x), \forall i \in I_2, \\ & 1 + \psi_{f_i}(\gamma, \alpha_i, x) \leq 0, \forall i \in [I] \setminus I_2, \\ & \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_i \geq 0, \forall i \in [I] \setminus I_2. \end{aligned} \right\}$$

Note that for each $i \in I$, $\psi_{f_i}(\gamma, \alpha_i, x) = -\alpha_i w_i(x) + \bar{\psi}_{-h_i}(\gamma, \alpha_i, x)$. Since function h_i is positive homogeneous in ξ and irrelevant with x , by Lemma 9, $\bar{\psi}_{-h_i}(\gamma, \alpha_i, x)$ is convex in (γ, α_i) and hence its domain is. Hence, we arrive at (5.41). \square

Note that (5.41b) is a convex constraint for each $i \in [I]$ and is second order cone representable by introducing a new variable $0 \leq q_i \leq w_i(x)$. Then (5.41b) is equivalent to

$$2^2 + (\alpha_i - q_i)^2 \leq (\alpha_i + q_i)^2, 0 \leq q_i \leq w_i(x), \forall i \in [I].$$

Note that Theorem 2 in [47] is a special case of Theorem 26, where in [47], it is assumed that $w_i(x)$ is an affine function and $h_i(\xi)$ is a linear function for each $i \in [I]$.

The following example demonstrates an application of Theorem 26.

Example 7. Consider a stochastic lot-sizing problem. There are I time periods and at each time period $i \in [I]$, ξ_i represents the random demand and c_i, f_i are the production cost and fixed cost, respectively. The production for each time period cannot exceed M . There are two types of decision variables, x_i represents production level and y_i represents production set up at time i , i.e., $y_i = 1$ if $x_i > 0$; 0, otherwise. Suppose that we know the mean of the demand at each period and the total deviation of demand, thus \mathcal{P} is defined as

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\tilde{\xi}] = 0, \mathbb{E}_{\mathbb{P}}[\|\tilde{\xi}\|_q] \leq g \right\},$$

where $\Xi = \mathbb{R}^I$ and $\xi = \tilde{\xi} + \mu$.

Now the entire distributionally robust lot-sizing problem is formulated as

$$v^* = \min \quad c^\top x + f^\top y,$$

$$\text{s.t.} \quad x_i \leq M y_i, \forall i \in [I], \tag{5.43a}$$

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left[\sum_{j=1}^i x_j \geq \sum_{j=1}^i (\tilde{\xi}_j + \mu_j), \forall i \in [I] \right] \geq 1 - \epsilon, \tag{5.43b}$$

$$y \in \{0, 1\}^I, x \geq 0. \tag{5.43c}$$

Let us define matrix $A \in \mathbb{R}^{I \times I}$ as $A(i, j) = 1$ if $j \leq i$; 0, otherwise; and a_i denote i th row of A . From Proposition 29, we have $w_i(x) = a_i \cdot (x - \mu)$, $h_i(\xi) = a_i \cdot \tilde{\xi}$, thus $\bar{\psi}_0(\gamma) = \|\gamma_1\|_{\frac{q}{q-1}} - \gamma_2$, $\bar{\psi}_{-h_i}(\gamma, -\alpha_i) = \|\gamma_1 + \alpha_i a_i\|_{\frac{q}{q-1}} - \gamma_2$ in Theorem 26. It is easy to see that

$$I_2 = \left\{ i \in [I] : \sup_{\tilde{\xi} \in \mathbb{R}^I} \left[\sum_{j=1}^i \tilde{\xi}_j \right] \leq 0 \right\} = \emptyset.$$

Thus, set Z is reformulated as

$$Z_C = \left\{ x : \begin{array}{l} g\gamma_2 \leq \epsilon, \\ 1 \leq \alpha_i a_i \cdot (x - \mu), \forall i \in [I], \\ \|\gamma_1\|_{\frac{q}{q-1}} \leq \gamma_2, \|\gamma_1 + \alpha_i a_i\|_{\frac{q}{q-1}} \leq \gamma_2, \forall i \in [I], \\ \gamma_2 \geq 0, \alpha_i \geq 0, \forall i \in [I]. \end{array} \right\} \quad \begin{array}{l} (5.44a) \\ (5.44b) \\ (5.44c) \\ (5.44d) \end{array}$$

In above formulation, note that the larger γ_2 value implies a larger feasible region. Thus, at optimality, we must have $\gamma_2 = \frac{\epsilon}{g}$. Therefore, Problem (5.43) is equivalent to the following mixed integer convex program:

$$v^* = \min \quad c^\top x + f^\top y, \quad (5.45a)$$

$$\text{s.t.} \quad x_i \leq M y_i, \forall i \in [I], \quad (5.45b)$$

$$1 \leq \alpha_i a_i \cdot (x - \mu), \forall i \in [I], \quad (5.45c)$$

$$\|\gamma_1\|_{\frac{q}{q-1}} \leq \frac{\epsilon}{g}, \|\gamma_1 + \alpha_i a_i\|_{\frac{q}{q-1}} \leq \frac{\epsilon}{g}, \forall i \in [I], \quad (5.45d)$$

$$\alpha_i \geq 0, \forall i \in [I], y \in \{0, 1\}^I, x \in \mathbb{R}_+^I. \quad (5.45e)$$

Note that variables α in (5.45) can be interpreted as a safety buffer, which guarantee that the inequalities (5.45c) are robust.

5.5 Binary DRCCP

In this section, we consider the case of binary decision variables and general moment ambiguity set \mathcal{P} defined in (5.3), i.e., $S \subseteq \{0, 1\}^n$ and linear chance constraints. We will first derive a mixed integer convex reformulation and then present a numerical study. Please note that the convexity results in Section 5.4 also apply to binary DRCCP.

5.5.1 Mixed integer convex formulation

Proposition 33. Suppose $S \subseteq \{0, 1\}^n$, $f_i(x, \xi) = (A^i x + a^i)^\top \xi + B^i x + b^i$ for each $i \in [I]$ and α in (5.15) can be upper bounded by a vector M for any $x \in S$. Consider a convex set

$$\bar{Y}_C = \left\{ x : \begin{array}{l} \lambda + \sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} g_t \gamma_t \geq 1 - \epsilon, \\ \lambda + \psi_0(\gamma) \leq 1, \\ \lambda + \psi_{\bar{f}_i}(\gamma, 1, (\alpha_i, y_i)) \leq 0, \forall i \in [I], \\ 0 \leq y_j^i \leq M_i x_j, \alpha_i - M_i(1 - x_j) \leq y_j^i \leq \alpha_i, \forall i \in [I], j \in [n], \\ \gamma_t \geq 0, \forall t \in \mathcal{T}_2, \alpha_i \geq 0, \forall i \in [I], \end{array} \right\} \quad \begin{array}{l} (5.46a) \\ (5.46b) \\ (5.46c) \\ (5.46d) \\ (5.46e) \end{array}$$

where for each $i \in [I]$, $\bar{f}_i(\alpha_i, y_i) = (A^i y_i + a^i \alpha_i)^\top \xi + B^i y_i + b^i \alpha_i$, then

$$S \cap \bar{Y}_C = S \cap Y_C \subseteq S \cap Z.$$

Proof. When $f_i(x, \xi) = (A^i x + a^i)^\top \xi + B^i x + b^i$ for each $i \in [I]$, by Definition 8, we have

$$\psi_{f_i}(\gamma, \alpha_i, x) = \sup_{\xi \in \Xi} \left(\sum_{t \in \mathcal{T}_1 \cup \mathcal{T}_2} \phi_t(\xi) \gamma_t - \alpha_i ((A^i x + a^i)^\top \xi + B^i x + b^i) \right).$$

Since $\alpha_i \leq M_i$ for each $i \in [I]$, let us define new variables y such that $y^i = \alpha_i x$, which can be linearized via McCormick inequalities [71] as $0 \leq y_j^i \leq M_i x_j, \alpha_i - M_i(1 - x_j) \leq y_j^i \leq \alpha_i$, i.e., (5.46d). This linearization is exact for any $x \in \{0, 1\}^n$. Thus, $S \cap \bar{Y}_C = S \cap Y_C$. \square

Proposition 33 tells that to optimize over $S \cap Y_C$ is equivalent to optimize over $S \cap \bar{Y}_C$, which is a mixed integer convex set instead of a mixed integer nonconvex set. Above we have assumed the existence of vector M and one sufficient condition is Corollary 7. Moreover, the next theorem shows that under some other cases, the variables α in (5.46) have closed-form bounds.

Theorem 27. Suppose the ambiguity set be defined as

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\xi_i] = \mu_i, \mathbb{E}_{\mathbb{P}}[(\xi_i - \mu_i)(\xi_i - \mu_i)^\top] \preceq \Sigma_i, \forall i \in [I] \right\},$$

where $\Xi = \mathbb{R}^{n \times I}$, and $\Sigma_i \succ 0$ for each $i \in [I]$. Suppose $S \subseteq \{0, 1\}^n$, $f_i(x, \xi) = (Ax + a)^\top \xi_i + B^i x + b^i$ for each $i \in [I]$, then

$$S \cap (X_C \cup \bar{Y}_C) = S \cap Z.$$

Sets X_C and \bar{Y}_C in (5.46) are defined as

$$X_C = \{x : Ax + a = 0, B^i x + b^i \geq 0, \forall i \in [I]\}, \quad (5.47)$$

and

$$\bar{Y}_C = \left\{ x : \begin{aligned} & \lambda - \sum_{j \in [I]} \left[\left\langle A^\top \Sigma_j A, w_{..j} \right\rangle + 2a^\top \Sigma_j A z_{.j} + a^\top \Sigma_j a \gamma_{2j} \right] \geq 1 - \epsilon, \\ & \lambda + \sum_{j \in [I]} t_{0j} \leq 1, \\ & \lambda + \sum_{j \in [I]} t_{ij} \leq \alpha_i \left((Ax + a)^\top \mu_i + b^i \right) + B^i y^i, \forall i \in [I], \\ & \gamma_{1j}^2 \leq 4t_{ij}\gamma_{2j}, \forall i \in [I], j \in [I] \setminus \{i\}, \\ & (\gamma_{1i} - \alpha_i)^2 \leq 4t_{ii}\gamma_{2i}, \gamma_{1i}^2 \leq 4t_{0i}\gamma_{2i}, \forall i \in [I], \\ & 0 \leq y_j^i \leq M_i x_j, \alpha_i - M_i(1 - x_j) \leq y_j^i \leq \alpha_i, \forall i \in [I], j \in [n], \\ & 0 \leq z_{ij} \leq \frac{\epsilon}{\underline{\delta}\eta} x_i, \gamma_{2j} - \frac{\epsilon}{\underline{\delta}\eta} (1 - x_i) \leq z_{ij} \leq \gamma_{2j}, \forall i, j \in [I], \\ & 0 \leq w_{ikj} \leq \frac{\epsilon}{\underline{\delta}\eta} x_i, 0 \leq w_{ikj} \leq \frac{\epsilon}{\underline{\delta}\eta} x_k, \\ & \gamma_{2j} - \frac{\epsilon}{\underline{\delta}\eta} (2 - x_i - x_j) \leq w_{ikj} \leq \gamma_{2j}, \forall i, j, k \in [I], \\ & \gamma_{2i} \geq 0, \alpha_i \geq 0, \forall i \in [I]. \end{aligned} \right\} \quad (5.48)$$

with

$$M_i = \frac{4\epsilon}{\underline{\delta}\eta} \left[(b^i + \mu_i^\top a^i + \|B^i + \mu_i^\top A^i\|_1) + \sqrt{(b^i + \mu_i^\top a^i + \|B^i + \mu_i^\top A^i\|_1)^2 + \underline{\delta}\eta - \frac{\underline{\delta}\eta}{2\epsilon}} \right]$$

for each $i \in [I]$, where $\eta = \min_{x \in \{0,1\}^n: Ax+a \neq 0} \|Ax + a\|_2^2$, $\underline{\delta}$ is the smallest eigenvalue of matrices $\{\Sigma_j\}_{j \in [I]}$, and $w_{..j}$ denotes the matrix (w_{ikj}) for each j and $z_{.j}$ denotes the vector (z_{ij}) for each j .

Proof. We will separate the proof into four parts.

- (i) Suppose for any $x \in \{0, 1\}^n$ such that $Ax + a = 0$, then by (5.1c), we must have $B^i x + b^i \geq 0$ for all $i \in [I]$. Hence, this implies that $x \in X_C$. Also note that for any $x \in S \cap Z \setminus X_C$, we must have $I_1(x) = [I]$, i.e., $x \in S \cap Y_C$. Therefore, $S \cap (X_C \cup Y_C) = S \cap Z$.

It remains to show the existence of vector M such that $\alpha_i \leq M_i$ for each $i \in [I]$ and $x \in S \cap Z \setminus X_C$.

- (ii) From now on, we assume that $\|Ax + a\|_2 \neq 0$ for all $x \in S \cap Y_C$. Define $\xi_i = \tilde{\xi}_i + \mu_i$ for each $i \in [I]$. Then the ambiguity set is equivalent to

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\tilde{\xi}_i] = 0, \mathbb{E}_{\mathbb{P}}[\tilde{\xi}_i \tilde{\xi}_i^\top] \preceq \Sigma_i, \forall i \in [I] \right\},$$

where $\Xi = \mathbb{R}^{n \times I}$. Also, the uncertain constraint is $f_i(x, \tilde{\xi}_i) = (Ax + a)^\top \tilde{\xi}_i + (Ax + a)^\top \mu_i + B^i x + b^i$ for each $i \in [I]$.

For any given $x \in Y_C$, replacing $\tilde{\xi}_i$ by $\zeta_i = (Ax + a)^\top \tilde{\xi}_i$ and by the standard random variable changing (c.f. [36]) and Theorem 1 in [79], the ambiguity set can be further replaced as

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I) : \mathbb{E}_{\mathbb{P}}[\zeta_i] = 0, \mathbb{E}_{\mathbb{P}}[\zeta_i^2] \leq (Ax + a)^\top \Sigma_i (Ax + a), \forall i \in [I] \right\},$$

with the uncertain constraint $f_i(x, \zeta_i) = \zeta_i + (Ax + a)^\top \mu_i + B^i x + b^i$ for each $i \in [I]$.

- (iii) Next by Definition 8, we have

$$\begin{aligned} \psi_{f_i}(\gamma, \alpha_i, x) &= \sum_{j \in [I], j \neq i} \sup_{\zeta_j \in \mathbb{R}} [\gamma_{1j} \zeta_j - \gamma_{2j} \zeta_j^2] \\ &\quad + \sup_{\zeta_i \in \mathbb{R}} \left[\gamma_{1i} \zeta_i - \gamma_{2i} \zeta_i^2 - \alpha_i \left(\zeta_i + (Ax + a)^\top \mu_i + B^i x + b^i \right) \right] \\ &= \min \left\{ \sum_{j \in [I]} t_{ij} - \alpha_i \left((Ax + a)^\top \mu_i + B^i x + b^i \right) : \gamma_{1j}^2 \leq 4t_{ij}\gamma_{2j}, \forall j \in [I] \setminus \{i\}, \right. \\ &\quad \left. (\gamma_{1i} - \alpha_i)^2 \leq 4t_{ii}\gamma_{2i} \right\} \end{aligned}$$

where the second equality is due to 1-dimensional S-Lemma in [12]. Similarly,

$$\psi_0(\gamma) = \min \left\{ \sum_{j \in [I]} t_{0j} : \gamma_{1j}^2 \leq 4t_{0j}\gamma_{2j}, \forall j \in [I] \right\}.$$

Then by replacing minimum operator with its equivalent “existence” argument, set Y_C can be formulated as

$$Y_C = \left\{ x : \begin{array}{l} \lambda - \sum_{j \in [I]} \gamma_{2j} (Ax + a)^\top \Sigma_j (Ax + a) \geq 1 - \epsilon, \\ \lambda + \sum_{j \in [I]} t_{0j} \leq 1, \\ \lambda + \sum_{j \in [I]} t_{ij} \leq \alpha_i \left((Ax + a)^\top \mu_i + B^i x + b^i \right), \forall i \in [I], \\ \gamma_{1j}^2 \leq 4t_{ij}\gamma_{2j}, \forall i \in [I], j \in [I] \setminus \{i\}, \\ (\gamma_{1i} - \alpha_i)^2 \leq 4t_{ii}\gamma_{2i}, \forall i \in [I], \\ \gamma_{1j}^2 \leq 4t_{0j}\gamma_{2j}, \forall j \in [I], \\ \gamma_{2i} \geq 0, \alpha_i \geq 0, \forall i \in [I]. \end{array} \right\} \quad \begin{array}{l} (5.49a) \\ (5.49b) \\ (5.49c) \\ (5.49d) \\ (5.49e) \\ (5.49f) \\ (5.49g) \end{array}$$

(iv) Now we show the existence of upper bounds on $\{\alpha_i\}_{i \in [I]}$ and $\{\gamma_{2i}\}_{i \in [I]}$ with following four steps.

(a) First of all, we observe that $t_{0j} \geq 0$ for each $j \in [I]$; otherwise, it contradicts that (5.49f). Thus, (5.49b) implies that $\lambda \leq 1$ and hence $\sum_{j \in [I]} \gamma_{2j} (Ax + a)^\top \Sigma_j (Ax + a) \leq \epsilon$.

Let $\underline{\delta}$ be the smallest eigenvalue of matrices $\{\Sigma_j\}_{j \in [I]}$. Hence, $\Sigma_j \succeq \underline{\delta} I_e$ for each $j \in [I]$, where I_e is the identity matrix.

Clearly, $\underline{\delta} > 0$ since $\Sigma_j \succ 0$ for each $j \in [I]$. Thus, $\sum_{j \in [I]} \gamma_{2j} (Ax + a)^\top \Sigma_j (Ax + a) \leq \epsilon$ and $\Sigma_j \succeq \underline{\delta} I$ for each $j \in [I]$ imply that

$$\underline{\delta} \|Ax + a\|_2^2 \sum_{j \in [I]} \gamma_{2j} \leq \epsilon.$$

Let $\eta = \min_{x \in \{0,1\}^n: Ax+a \neq 0} \|Ax + a\|_2^2$, thus γ_{2j} is bounded by $\frac{\epsilon}{\underline{\delta}\eta}$.

(b) Therefore, (5.49d), (5.49e) and (5.49f) can be relaxed by replacing $\{\gamma_{2j}\}_{j \in [I]}$ with their lower bound $\frac{\epsilon}{\underline{\delta}\eta}$ as below

$$t_{ij} \geq \frac{\underline{\delta}\eta}{4\epsilon} \gamma_{1j}^2, \forall i \in [I], j \in [I] \setminus \{i\}, \quad (5.50)$$

$$t_{ii} \geq \frac{\underline{\delta}\eta}{4\epsilon} (\gamma_{1i} - \alpha_i)^2, \forall i \in [I], \quad (5.51)$$

$$t_{0j} \geq \frac{\delta\eta}{4\epsilon} \gamma_{1j}^2, \forall j \in [I]. \quad (5.52)$$

(c) Since $t_{ij}, t_{0j} \geq 0$ for all $i \in [I], j \in [I] \setminus \{i\}$ and $\Sigma_j \succ 0$ for all $j \in [I]$, thus (5.49a), (5.49b) and (5.49c) imply that

$$\begin{aligned} \lambda &\geq 1 - \epsilon, \\ \lambda + t_{ii} &\leq \alpha_i [b^i + \mu_i^\top a + (B^i + \mu_i^\top A)x], \forall i \in [I], \\ \lambda + t_{0i} &\leq 1, \forall i \in [I]. \end{aligned}$$

Together with (5.51) and (5.52), these above inequalities are further reduced to

$$\begin{aligned} -2\lambda &\leq -2(1 - \epsilon), \\ \lambda + \frac{\delta\eta}{4\epsilon} (\gamma_{1i} - \alpha_i)^2 - \alpha_i [b^i + \mu_i^\top a + (B^i + \mu_i^\top A)x] &\leq 0, \forall i \in [I], \\ \lambda + \frac{\delta\eta}{4\epsilon} \gamma_{1i}^2 &\leq 1, \forall i \in [I]. \end{aligned}$$

Summing these inequalities up for each $i \in [I]$ yields

$$\frac{\delta\eta}{4\epsilon} [(\gamma_{1i} - \alpha_i)^2 + \gamma_{1i}^2] - \alpha_i [b^i + \mu_i^\top a + (B^i + \mu_i^\top A)x] \leq 2\epsilon - 1.$$

(d) Using the fact that $(r + s)^2 + s^2 \geq \frac{r^2}{2}$, we have

$$\frac{\delta\eta}{8\epsilon} \alpha_i^2 - \alpha_i [b^i + \mu_i^\top a + (B^i + \mu_i^\top A)x] \leq 2\epsilon - 1, \forall i \in [I].$$

As $x \in \{0, 1\}^n$, we have $(B^i + \mu_i^\top A)x \leq \|B^i + \mu_i^\top A\|_1$. Thus we arrive at the following inequality

$$\frac{\delta\eta}{8\epsilon} \alpha_i^2 - \alpha_i (b^i + \mu_i^\top a + \|B^i + \mu_i^\top A\|_1) + 1 - 2\epsilon \leq 0, \forall i \in [I].$$

Hence, α can be upper bounded by

$$\begin{aligned} M_i &= \frac{4\epsilon}{\delta\eta} \left[(b^i + \mu_i^\top a + \|B^i + \mu_i^\top A\|_1) \right. \\ &\quad \left. + \sqrt{(b^i + \mu_i^\top a + \|B^i + \mu_i^\top A\|_1)^2 + \delta\eta - \frac{\delta\eta}{2\epsilon}} \right] \end{aligned}$$

for each $i \in [I]$.

(v) As $\alpha_i \leq M_i$ and $x \in \{0, 1\}^n$ for each $i \in [I]$, let us define new variables y such that $y^i = \alpha_i x$, which can be linearized via McCormick inequalities as

$$0 \leq y_j^i \leq M_i x_j, \alpha_i - M_i(1 - x_j) \leq y_j^i \leq \alpha_i.$$

Also since $\gamma_{2j} \leq \frac{\epsilon}{\underline{\delta}\eta}$, thus, let $z_{ij} = \gamma_{2j} x_i, w_{ikj} = \gamma_{2j} x_i x_k$ for all $i, k, j \in [I]$, which also can be linearized via McCormick inequalities as

$$\begin{aligned} 0 \leq z_{ij} &\leq \frac{\epsilon}{\underline{\delta}\eta} x_i, \gamma_{2j} - \frac{\epsilon}{\underline{\delta}\eta} (1 - x_i) \leq z_{ij} \leq \gamma_{2j}, \\ 0 \leq w_{ikj} &\leq \frac{\epsilon}{\underline{\delta}\eta} x_i, 0 \leq w_{ikj} \leq \frac{\epsilon}{\underline{\delta}\eta} x_k, \gamma_{2j} - \frac{\epsilon}{\underline{\delta}\eta} (2 - x_i - x_j) \leq w_{ikj} \leq \gamma_{2j}. \end{aligned}$$

Thus, we arrive at (5.48).

□

The following example illustrates an application of Theorem 27. This example has been studied in [32], where the authors presented several different heuristic (approximate) algorithms. Instead, we show that the feasible region of this problem can be approximated almost exactly as a mixed integer second order conic program (SOCP). Thus, any mixed integer SOCP approach could be used to solve it.

Example 8. (multi-dimensional binary knapsack problem) Consider a variant of Example 14 where $x \in \{0, 1\}^n$, i.e. $x_j = 1$ if j th item being picked, 0 otherwise. Suppose that we know the mean of weight vector of each knapsack and its second moment, thus \mathcal{P} is defined as

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) : \mathbb{E}_{\mathbb{P}}[\tilde{\xi}_i] = 0, \mathbb{E}_{\mathbb{P}}[\tilde{\xi}_i \tilde{\xi}_i^\top] \preceq \Sigma_i, \forall i \in [I] \right\},$$

where $\Xi = \mathbb{R}^{n \times I}$ and $\xi_i = \tilde{\xi}_i + \mu_i$. Without loss of generality, we assume that $\mu_i \geq 0, \Sigma_i \succ 0$ for each $i \in [I]$.

Now the entire distributionally robust multi-dimensional knapsack problem is formulated as

$$v^* = \max \quad c^\top x,$$

$$\text{s.t. } x \in \{0, 1\}^n,$$

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[F(x, \tilde{\xi}) \geq 0] \geq 1 - \epsilon,$$

where $f_i(x, \tilde{\xi}) = b_i - x^\top \mu_i - x^\top \tilde{\xi}_i$ for each $i \in [I]$.

By Theorem 27, we must have $S \cap Z = S \cap (X_C \cup \bar{Y}_C)$, where $X_C = \{0\}$ and \bar{Y}_C is equivalent to

$$\bar{Y}_C = \left\{ x : \begin{aligned} & \lambda - \sum_{j \in [I]} \langle \Sigma_j, w_{..j} \rangle \geq 1 - \epsilon, \\ & \lambda + \sum_{j \in [I]} t_{0j} \leq 1, \\ & \lambda + \sum_{j \in [I]} t_{ij} \leq \alpha_i b_i - \mu_i^\top y^i, \forall i \in [I], \\ & \gamma_{1j}^2 \leq 4t_{ij}\gamma_{2j}, \forall i \in [I], j \in [I] \setminus \{i\}, \\ & (\gamma_{1i} - \alpha_i)^2 \leq 4t_{ii}\gamma_{2i}, \forall i \in [I], \\ & \gamma_{1j}^2 \leq 4t_{0j}\gamma_{2j}, \forall j \in [I], \\ & 0 \leq y_j^i \leq M_i x_j, \alpha_i - M_i(1 - x_j) \leq y_j^i \leq \alpha_i, \forall i \in [I], j \in [n], \\ & 0 \leq w_{ikj} \leq \frac{\epsilon}{\underline{\delta}\eta} x_i, 0 \leq w_{ikj} \leq \frac{\epsilon}{\underline{\delta}\eta} x_k, \\ & \gamma_{2j} - \frac{\epsilon}{\underline{\delta}\eta}(2 - x_i - x_j) \leq w_{ikj} \leq \gamma_{2j}, \forall i, j, k \in [I] \\ & \gamma_{2i} \geq 0, \alpha_i \geq 0, \forall i \in [I]. \end{aligned} \right\} \quad (5.53)$$

where

$$M_i = \frac{4\epsilon}{\underline{\delta}\eta} \left[(b^i + \|\mu_i\|_1) + \sqrt{(b^i + \|\mu_i\|_1)^2 + \underline{\delta}\eta - \frac{\underline{\delta}\eta}{2\epsilon}} \right]$$

for each $i \in [I]$ with $\eta = \min_{x \in \{0,1\}^n: x \neq 0} \|x\|_2^2 = 1$ and $\underline{\delta}$ the smallest eigenvalue of matrices $\{\Sigma_j\}_{j \in [I]}$.

5.5.2 Numerical illustration

In this section, we present a numerical study to illustrate the strength of proposed formulation (5.53) corresponding to the multidimensional knapsack problem in Example 8. The instances are constructed from the problem set *mk-20-10* in [101]. The instances in this set are named *1-4-multi-N-i*, where $N \in \{100, 500, 1000, 3000\}$ denotes sample size of the

weight vector and there are 5 different instances for each sample size (i.e., $i \in \{1, 2, 3, 4, 5\}$). Each instance has 20 decision variables and 10 knapsack constraints, i.e, $n = 20, I = 10$. We compute μ, σ of the weight vector for each knapsack as the sample mean and covariance from the provided data.

Our first approach is to solve the mixed integer SOCP (5.53) exactly. We notice that the explicit upper bounds of α, γ_2 could be quite loose. Hence, instead, we enhance these bounds by maximizing these variables over the continuous relaxation of (5.53).

We compare our approach with the heuristic one proposed in [32]. The authors formulate to maximize $c^\top x$ over set $Z_C \cap \{0, 1\}^n$ in (5.13) as the following mixed integer nonconvex program

$$\begin{aligned}
& \max_{x \in \{0, 1\}^n} c^\top x, \\
& \text{s.t. } \lambda - \sum_{j \in [I]} \langle \Sigma_j, \gamma_{2j} \rangle \geq (1 - \epsilon)\beta, \\
& \lambda + \sum_{j \in [I]} t_{0j} \leq \beta, \\
& \lambda + \sum_{j \in [I]} t_{ij} \leq \alpha_i(b_i - \mu_i^\top x), \forall i \in [I], \\
& \begin{bmatrix} t_{ij} & -\frac{1}{2}\gamma_{1j}^\top \\ -\frac{1}{2}\gamma_{1j} & \gamma_{2j} \end{bmatrix} \succeq 0, \forall i \in [I], j \in [I] \setminus \{i\}, \\
& \begin{bmatrix} t_{ii} & -\frac{1}{2}(\gamma_{1i} + \alpha_i x)^\top \\ -\frac{1}{2}(\gamma_{1i} + \alpha_i x) & \gamma_{2i} \end{bmatrix} \succeq 0, \forall i \in [I], \\
& \begin{bmatrix} t_{0j} & -\frac{1}{2}\gamma_{1j}^\top \\ -\frac{1}{2}\gamma_{1j} & \gamma_{2j} \end{bmatrix} \succeq 0, \forall j \in [I], \\
& \gamma_{2i} \succeq 0, \alpha_i > 0, \forall i \in [I].
\end{aligned} \tag{5.54}$$

The solution approach is summarized below. First of all, given an α , solve the continuous relaxation of (5.54) with α fixed, which is an SDP. Let \hat{x} be the corresponding optimal solution. Then, fix $x = \hat{x}$, change the objective function to $\max \lambda - \sum_{j \in [I]} \langle \Sigma_j, \gamma_{2j} \rangle$ (i.e.,

maximize the largest probability) and solve the corresponding continuous relaxation problem with optimal solution $\hat{\alpha}$. In the next step, let $\alpha = \hat{\alpha}$, and iterate. This procedure terminates whenever the values of α and x no longer change. Suppose, at the end of this procedure, $\alpha = \hat{\alpha}, x = \hat{x}$. In general, $\hat{x} \in [0, 1]^n$ is not binary. So the final step is randomized rounding, i.e. treat $x_i \in [0, 1]$ as a Bernoulli random variable with probability \hat{x}_i for each $i \in [n]$, generate a sample \tilde{x} , then check the feasibility of \tilde{x} to (5.54). This step could repeat multiple times until finding several candidate solutions (e.g., 5 solutions) and of course, choosing the best one as the output.

We use commercial solver CPLEX for the first approach, while CVX for the second approach. The results are listed in Table 8. We use $v_{\text{SOCP}}, t_{\text{SOCP}}$ to denote the objective value and the total running time of first approach (including big- M strengthening time), while use $v_{\text{H}}, t_{\text{H}}, \text{gap}$ for the objective value, the total running time and the optimality gap of the second approach (the heuristic one). All instances were executed on a laptop with a 2.67 GHz processor and 4GB RAM, while CPLEX 12.5.1 and CVX 2.1 were used with their default settings.

In Table 8, we observe that the solution time of both methods are in general quite similar, while on average, the exact approach (3463s) takes less time than the heuristic one (4268s). If we compare the solution quality, it can be seen that the solution of the heuristic method is quite unpredictable, i.e., for some instances, it finds a very good solution but for others, it does not. The average gap of the heuristic solutions is around 20%. On the other hand, the exact approach can find the optimal solution within an hour and a half for majority of instances (34 over 40 instances). These results demonstrate the effectiveness of the exact approach proposed in this chapter.

Table 8: Performance comparison of exact and the heuristic methods

ϵ	N	Index	Exact Approach (5.53)		Heuristic Approach by [32]		
			v_{SOCP}	t_{SOCP}	v_{H}	t_{H}	gap(%)
0.05	100	1	2925	2291	2860	3069	2.2%
		2	3950	8412	3750	2852	5.1%
		3	2910	3740	2580	3111	11.3%
		4	3430	2061	1900	3298	44.6%
		5	3860	3756	3510	12294	9.1%
0.1	100	1	3140	8945	2860	4972	8.9%
		2	4460	5546	4330	3850	2.9%
		3	4140	3914	2610	4537	37.0%
		4	4100	5597	4010	3462	2.2%
		5	4380	7429	3935	4698	10.2%
0.05	500	1	3480	3286	1710	4846	50.9%
		2	3400	3471	2000	4564	41.2%
		3	3800	2080	3680	3917	3.2%
		4	3050	4300	1690	3525	44.6%
		5	4010	3897	3980	3198	0.7%
0.1	500	1	4100	5304	2580	3207	37.1%
		2	4270	2188	2030	3194	52.5%
		3	4480	4269	4380	3506	2.2%
		4	4050	5197	3760	3376	7.2%
		5	4530	5870	4240	3350	6.4%
0.05	1000	1	4150	1430	2490	3431	40.0%
		2	3280	789	2580	5451	21.3%
		3	4060	1369	3920	3428	3.4%
		4	2640	1125	2170	3489	17.8%
		5	2560	1300	2090	4096	18.4%
0.1	1000	1	4560	2527	4550	4100	0.2%
		2	4280	2655	2610	4543	39.0%
		3	4560	4304	4360	3441	4.4%
		4	3500	1517	2490	3877	28.9%
		5	3150	3424	2490	3720	21.0%
0.05	3000	1	2905	2053	2530	5455	12.9%
		2	2770	1701	2330	4482	15.9%
		3	2840	2546	2610	3844	8.1%
		4	2860	2361	2240	3830	21.7%
		5	2770	3087	2290	7839	17.3%
0.1	3000	1	3680	2026	2170	3904	41.0%
		2	3460	1765	2170	3837	37.3%
		3	3610	4282	2580	7537	28.5%
		4	3560	4609	2490	4069	30.1%
		5	3470	2097	2170	3512	37.5%

CHAPTER VI

DISTRIBUTIONALLY ROBUST CHANCE CONSTRAINED OPTIMAL POWER FLOW WITH RENEWABLES: A CONIC REFORMULATION

6.1 *Introduction*

Recently with growing interests in environmentally friendly power generation such as wind, solar, geothermal energy [63, 96, 107], optimal power flow (OPF) under uncertainty has attracted much attention from researchers [9, 19, 102, 110, 125, 130, 131]. A particular issue caused by renewables is the voltage fluctuations which can lead to severe issues, for example, overloaded transmission lines [31, 69]. To mitigate these issues, [9, 19] proposed a chance constrained optimal power flow model (CC-OPF) that constrains the overloading probability. This chapter extends this work by enforcing power flow within lower and upper bounds simultaneously and show tractability results of such an approach under data driven distributionally robust setting.

There are many works on solving OPF, unit commitment (UC) problem or reserve scheduling via stochastic programming [78, 90, 109], robust optimization [17, 40, 49, 50, 111, 113, 122, 132], and chance constrained program approaches [18, 19, 60, 75, 84, 102, 110, 125] (see [65] for some discussions). Stochastic programming approaches highly rely on the underlying distribution, which could be unknown in many cases, and the performance of solution algorithms is usually very sensitive to the distribution used [94]. Robust optimization is often too conservative [11], while chance constrained programming is less conservative but is often NP-hard [67, 73]. Thus, to overcome the difficulties from these two approaches, here we adopt a distributionally robust chance constrained approach, which allows violation of uncertain constraints with a small probability for a large class of probability distributions and could be reformulated as a tractable convex program [9, 24, 118, 135].

There are two concerns about existing literatures on CC-OPF formulations. It is known

that each transmission line as well as each bus (node) in general has lower and upper bound limits. However, most works [9, 19, 65, 129] treat the lower- and upper- bound overloading separately, which is an inexact approximation. To the best of our knowledge, [64] is the only known work treating lower and upper bounds simultaneously. However the results in [64] highly depend on the assumption of a Gaussian distribution on the underlying uncertainties. In this chapter, we will consider incorporating power flow within lower and upper bounds simultaneously and our results are distribution-free.

Another issue regarding previous CC-OPF studies is that they assumed a particular distribution of renewables' output. For example, [19, 65, 83] assumed that the prior distributions of renewables are Gaussian while [62] assumed that it is Weibull. However, these assumptions might not be true in practice [9]. In general the underlying probability distributions of renewables are not known or are hard to estimate from empirical data. Thus, to hedge against the uncertainty of probability distributions we consider a data driven distributionally robust chance constrained optimal power flow model (DRCC-OPF) by considering the overload within the upper and lower bounds jointly with high probability. And the underlying probability distribution comes from a family of distributions (called "*ambiguity set*") that share the same mean and covariance matrix estimated from empirical data.

Distributionally robust chance constrained problems with multiple uncertain constraints (joint-DRCCP) are very challenging [46]. There are only few settings under which joint DRCCP can be equivalently reformulated into a convex program. For example, in [47], they assumed right-hand uncertainty with mean dispersion moment ambiguity set, and proposed convex reformulations. Recently, [118] explored several sufficient conditions under which joint DRCCP is convex and [121] showed that joint DRCCP with separable structure (the uncertainties are separable across the individual uncertain constraints and their corresponding distribution families) can be convex with small risk parameter. However none of known sufficient conditions for convexity can be directly applied to the two-sided DRCC-OPF here, where in two-sided chance constraint, the uncertain constraints are defined as lower and upper bounds of one uncertain affine function. This chapter fills the gap in [118] by showing that joint DRCCP with two-sided constraint has a conic reformulation.

The remaining of the chapter is organized as follows. Section 6.2 lists all the notations used in this chapter. Section 6.3 introduces the model formulation and Section 6.4 shows how to reformulate the model into a convex second order cone program. Section 6.5 numerically illustrates the strengths of the proposed model.

6.2 Nomenclature

6.2.1 Sets

\mathcal{V} = set of all buses

\mathcal{E} = set of all transmission lines linking two buses

\mathcal{G} = subset of buses that house generators

\mathcal{W} = subset of buses that holds uncertain power sources (wind farms)

6.2.2 Parameters

μ_j = average generation at bus $j \in \mathcal{W}$ ($\mu_j = 0$, for all $j \in \mathcal{V} \setminus \mathcal{W}$)

ω_j = random fluctuation of power generation at uncertain power source $j \in \mathcal{W}$ ($\omega_i = 0$, if $i \in \mathcal{V} \setminus \mathcal{W}$)

Σ = known covariance matrix of random vector ω

d_i = demand at bus $i \in \mathcal{V}$

β_{ij} = the line susceptance between $(i, j) \in \mathcal{E}$ ($\beta_{ij} = \beta_{ji}$ by symmetry)

B = weighted Laplacian matrix defined as

$$B(i, j) = \begin{cases} -\beta_{ij}, & \text{if } (i, j) \in \mathcal{E} \\ \sum_{k: (k, j) \in \mathcal{E}} \beta_{kj}, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

for each (i, j)

\hat{B} the submatrix of B by removing the last row and column

\check{B} = pseudo-inverse of B defined as

$$\check{B} = \begin{bmatrix} \hat{B}^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

$\check{B}^{\mathcal{W}} = |V| \times |W|$ submatrix of \check{B} where its columns are from \mathcal{W}

$\check{\mathcal{B}}_i^{\mathcal{W}} = i\text{th row of } \check{\mathcal{B}}^{\mathcal{W}}$

$e = \text{all one vector}$

$c_i, r_i = \text{quadratic cost coefficients of generator } i \in \mathcal{G}$

$\hat{e}_{ij} = \text{risk parameter of violating the capacity of transmission line } (i, j) \in \mathcal{E}$

$\bar{e}_i = \text{risk parameter of violating the capacity of bus } i \in \mathcal{G}$

$f_{ij}^{\max} = \text{max capacity of transmission line } (i, j) \in \mathcal{E}$

$p_i^{\min} = \text{generation lower bound of bus } i \in \mathcal{G}$

$p_i^{\max} = \text{generation upper bound of bus } i \in \mathcal{G}$

6.2.3 Decision Variables

$\bar{\theta}_j = \text{be the phase of bus } j \in \mathcal{V}$

$\bar{p}_i = \text{regular generation at generator } i \in \mathcal{G} \ (\bar{p}_i = 0, \text{ if } i \in \mathcal{V} \setminus \mathcal{G})$

$\alpha_i = i\text{th assignment of total renewables to generator } i \in \mathcal{G} \ (\alpha_i = 0, \text{ if } i \in \mathcal{V} \setminus \mathcal{G})$

6.3 Model Formulation

In this section, we consider an extension of distributionally robust chance constrained optimal flow model (DRCC-OPF) proposed in [19].

In the optimal power flow problem, we suppose that there is a subset \mathcal{W} of the buses with uncertain power sources (e.g., wind farms). For each $j \in \mathcal{W}$, we model the uncertain power generated by $\mu_j + \omega_j$, where μ_j represents the mean of uncertain power generation and ω_j is a random variable with zero mean and covariance matrix denoted by Σ . The net output of bus $i \in \mathcal{G}$ is fluctuated by the output of wind generators. Let α_i for each $i \in \mathcal{G}$ be the proportion of wind power allocated to bus i , i.e., the output of bus $i \in \mathcal{G}$ is $\bar{p}_i - (e^\top \omega) \alpha_i$ with nonnegative variables \bar{p}_i, α_i and $\sum_{i \in \mathcal{G}} \alpha_i = 1$. Each bus $i \in \mathcal{V}$ has demand d_i . For notational convenience, we extend vectors $\omega, \mu, \alpha, \bar{p}$ to $\mathbb{R}^{|\mathcal{V}|}$ by letting $\omega_j = 0, \mu_j = 0$ for each $j \in \mathcal{V} \setminus \mathcal{W}$ and $\alpha_i = 0, \bar{p}_i = 0$ for each $i \in \mathcal{V} \setminus \mathcal{G}$.

Let $\bar{\theta}$ be the phases of all the buses. To approximate nonconvex AC power flow equations, we use DC-approximation. Thus, the power flow between line (i, j) is approximated as $\beta_{ij}(\bar{\theta}_i - \bar{\theta}_j)$ where $\beta_{ij} = \beta_{ji}$ denotes the line susceptance.

Following [19], a distributionally robust chance constrained optimal power flow problem

(DRCC-OPF) is formulated as

$$v^* = \min_{\bar{p}, \alpha, \bar{\theta}} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\sum_{i \in \mathcal{G}} c_i (\bar{p}_i + \alpha_i (e^\top \boldsymbol{\omega}) + r_i)^2 \right] \quad (6.1a)$$

$$\text{s.t. } \sum_{i \in \mathcal{G}} \alpha_i = 1, \quad (6.1b)$$

$$\sum_{i \in \mathcal{V}} (\bar{p}_i + \mu_i - d_i) = 0, \quad (6.1c)$$

$$B\bar{\theta} = \bar{p} + \mu - d, \quad (6.1d)$$

$$\begin{aligned} & \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\{\boldsymbol{\omega} : |\beta_{ij}(\bar{\theta}_i - \bar{\theta}_j) + [\check{\mathcal{B}}^{\mathcal{W}}(\boldsymbol{\omega} - (e^\top \boldsymbol{\omega})\alpha)]_i \\ & - [\check{\mathcal{B}}^{\mathcal{W}}(\boldsymbol{\omega} - (e^\top \boldsymbol{\omega})\alpha)]_j| \leq f_{ij}^{\max}\} \geq 1 - \hat{\epsilon}_{ij}, \\ & \forall (i, j) \in \mathcal{E}, \end{aligned} \quad (6.1e)$$

$$\begin{aligned} & \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\{\boldsymbol{\omega} : p_i^{\min} \leq \bar{p}_i - (e^\top \boldsymbol{\omega})\alpha_i \leq p_i^{\max}\} \\ & \geq 1 - \bar{\epsilon}_i, \forall i \in \mathcal{G}, \end{aligned} \quad (6.1f)$$

$$\bar{p} \geq 0, \alpha \geq 0, \bar{p}_i = 0, \alpha_i = 0, \forall i \in \mathcal{V} \setminus \mathcal{G}. \quad (6.1g)$$

where (6.1a) is to optimize cost function where $c > 0$ and $r \in \mathbb{R}^{|\mathcal{G}|}$ are constant, (6.1b) implies that the total assignment of power from wind is 1, (6.1c) means on average, the total generation equals to the total demand, (6.1d) is the DC-approximation equation with

$$B(i, j) = \begin{cases} -\beta_{ij}, & \text{if } (i, j) \in \mathcal{E} \\ \sum_{k: (k, j) \in \mathcal{E}} \beta_{kj}, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

for each (i, j) , (6.1e) enforce that the worst case probability that the absolute flow on (i, j) does not exceed the maximum capacity f_{ij}^{\max} should be no smaller than $1 - \hat{\epsilon}_{ij}$ with pseudo-inverse of B

$$\check{B} = \begin{bmatrix} \hat{B}^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

and its submatrix $\check{\mathcal{B}}^{\mathcal{W}}$ and \hat{B} the submatrix of B by removing the last row and column; and (6.1f) ensures that with probability at least $1 - \bar{\epsilon}_i$, the generated power at i satisfied the lower bound p_i^{\min} and the upper bound p_i^{\max} , (6.1g) defines the boundary of variables. Here we assume the all the risk parameters are within $(0, 1)$.

Similar to [33, 51, 135], let us consider the ambiguity set defined by first and second moments as

$$\mathcal{P} = \{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{|\mathcal{V}|}) : \mathbb{E}_{\mathbb{P}}[\boldsymbol{\omega}] = 0, \mathbb{E}_{\mathbb{P}}[\boldsymbol{\omega}\boldsymbol{\omega}^\top] = \Sigma\} \quad (6.2)$$

where $\mathcal{P}_0(\mathbb{R}^{|\mathcal{V}|})$ denotes the set of all of probability measures on $\mathbb{R}^{|\mathcal{V}|}$ with a sigma algebra \mathcal{F} , and $\Sigma \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ is a positive semi-definite matrix (i.e., $\Sigma \succeq 0$). We remark that various other works have studied different moment ambiguity sets, for example, mean-dispersion or mean deviation ambiguity set [47, 121], ambiguity set with known sum of variances and covariance [33, 128, 129], or ambiguity set with a bounded support [102]. The results for other types of ambiguity set might be different from the one of (6.2), however, the proof technique here is very general and may be applicable to these settings as well.

As we know the mean and covariance of $\boldsymbol{\omega}$, the cost function (6.1a) is equivalent to

$$\begin{aligned} & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\sum_{i \in \mathcal{G}} c_i (\bar{p}_i + \alpha_i (e^\top \boldsymbol{\omega}) + r_i)^2 \right] \\ &= \sum_{i \in \mathcal{G}} \left(c_i (\bar{p}_i + r_i)^2 + c_i \alpha_i^2 e^\top \Sigma e \right). \end{aligned} \quad (6.3)$$

Thus, apart from the chance constraints (6.1e) and (6.1f), the DRCC-OPF formulation (6.1) is a convex quadratic optimization problem.

6.4 Convex Reformulation of Chance Constraints (6.1e) and (6.1f)

In this section, we will develop a deterministic convex formulation of (6.1) by reformulating the chance constraints (6.1e) and (6.1f) into equivalent convex constraints.

To reformulate the chance constraints (6.1e) and (6.1f), let us first consider a generic distributionally robust chance constrained set defined as follows:

$$Z := \left\{ x : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[|a(x)^\top \boldsymbol{\omega} + b(x)| \leq T] \geq 1 - \epsilon \right\} \quad (6.4)$$

where $a(x), b(x)$ are affine mappings. This set is defined by a distributionally robust two-sided chance constraint.

Note that (6.1e) and (6.1f) are special cases of (6.4), where in (6.1e), we let $a(x) = \beta_{ij}(\check{\mathcal{B}}_i^\mathcal{V} - \alpha_i e - \check{\mathcal{B}}_j^\mathcal{V} + \alpha_j e)$, $b(x) = \beta_{ij}(\bar{\theta}_i - \bar{\theta}_j)$, $T = f_{ij}^{\max}$ and in (6.1f), we let $a(x) = -\alpha_i e$, $b(x) = \bar{p}_i - \frac{p_i^{\max} + p_i^{\min}}{2}$, $T = \frac{p_i^{\max} - p_i^{\min}}{2}$.

6.4.1 Approximation by Two Single-sided Chance Constraints

Recently, including [9, 19, 125, 131], many studies tried to approximate the two-side chance constrained set (6.4) by two single-sided chance constraints. In particular, let

$$Z_A(\alpha) = \left\{ x : \begin{array}{l} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[a(x)^\top \boldsymbol{\omega} + b(x) \leq T] \geq 1 - \alpha, \\ \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[a(x)^\top \boldsymbol{\omega} + b(x) \geq -T] \geq 1 - \alpha. \end{array} \right\} \quad (6.5a)$$

By choosing $\alpha \sim \epsilon$ existing works use $Z_A(\alpha)$ to approximate Z . It turns out that sets $Z_A(\epsilon)$ and $Z_A(\epsilon/2)$ are outer and inner approximations of set Z and can be formulated as second order cone programs (SOCP).

Theorem 28. *Suppose the ambiguity set \mathcal{P} is defined as in (6.2), then set Z_A is equivalent to the following SOCP*

$$Z_A(\alpha) = \left\{ x : \begin{array}{l} b(x) + \sqrt{\frac{1-\alpha}{\alpha}} \sqrt{a(x)^\top \Sigma a(x)} \leq T, \\ -b(x) + \sqrt{\frac{1-\alpha}{\alpha}} \sqrt{a(x)^\top \Sigma a(x)} \leq T. \end{array} \right\} \quad (6.6a)$$

and $Z_A(\epsilon/2) \subseteq Z \subseteq Z_A(\epsilon)$.

Proof. For $x \in Z$, clearly, we have

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[|a(x)^\top \boldsymbol{\omega} + b(x)| \leq T] \leq \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[a(x)^\top \boldsymbol{\omega} + b(x) \leq T]$$

and

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[|a(x)^\top \boldsymbol{\omega} + b(x)| \leq T] \leq \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}[a(x)^\top \boldsymbol{\omega} + b(x) \geq -T].$$

Clearly, $Z \subseteq Z_A(\epsilon)$.

The result that $Z_A(\epsilon/2) \subseteq Z$ follows by Bonferroni approximation of joint chance constrained set (c.f. [73]). The equivalent reformulation of $Z_A(\alpha)$ follows by Theorem 3.1 in [24]. \square

As discussed in the sequel the approximations offered by $Z_A(\alpha)$ could be very crude, especially when the risk parameter ϵ is modest. In the next subsection, we will explore an exact convex reformulation of the set Z .

6.4.2 Exact Reformulation

Our main result is the following theorem which provides a convex reformulation of the two-sided chance constrained set (6.4) as an SOCP.

Theorem 29. *Suppose the ambiguity set \mathcal{P} is defined in (6.2), then set Z is equivalent to the following convex SOCP (involving two additional variables):*

$$Z = \left\{ \begin{array}{l} y^2 + a(x)^\top \Sigma a(x) \leq \epsilon(T - \pi)^2, \\ x : |b(x)| \leq y + \pi, \\ T \geq \pi \geq 0, y \geq 0. \end{array} \right\} \quad \begin{array}{l} (6.7a) \\ (6.7b) \\ (6.7c) \end{array}$$

Proof. Observe that $|b(x)| \leq T$ for each $x \in Z$. This is because by choosing $\omega_1 = 0$ with probability $1 - \epsilon$, $\omega_{i+1} = \sqrt{\frac{|\mathcal{W}|}{\epsilon}} \lambda_i^{0.5} u_i$, $\omega_{i+|\mathcal{W}|+1} = -\sqrt{\frac{|\mathcal{W}|}{\epsilon}} \lambda_i^{0.5} u_i$ with probability $\frac{\epsilon}{2|\mathcal{W}|}$ for each $i \in \mathcal{W}$, where λ_i and u_i are i th eigenvalue and eigenvector of Σ . Then, under this particular construction, we have $|b(x)| \leq T$.

Lemma 11.

$$Z = Z_1 \cup Z_2 \cup Z_3$$

where

$$Z_1 = \left\{ x : \begin{array}{l} \sqrt{\frac{1}{\epsilon}} \sqrt{a(x)^\top \Sigma a(x) + b(x)^2} \leq T \\ |b(x)| \leq \epsilon T, \end{array} \right\}$$

$$Z_2 = \left\{ x : \begin{array}{l} b(x) + \sqrt{\frac{1-\epsilon}{\epsilon}} \sqrt{a(x)^\top \Sigma a(x)} \leq T, \\ b(x) \geq \epsilon T, \end{array} \right\}$$

and

$$Z_3 = \left\{ x : \begin{array}{l} -b(x) + \sqrt{\frac{1-\epsilon}{\epsilon}} \sqrt{a(x)^\top \Sigma a(x)} \leq T, \\ b(x) \leq -\epsilon T. \end{array} \right\}$$

Proof. Suppose $x \in Z$, then by the standard random variable transformation (c.f. [36]) and Theorem 1 in [79] (see also in [59, 118]), set Z is equivalent to

$$Z := \left\{ x : \inf_{\mathbb{P} \in \mathcal{P}_1} \mathbb{P}[|\xi| \leq T] \geq 1 - \epsilon \right\}$$

where

$$\mathcal{P}_1 = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}) : \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] = b(x), \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}^2] = a(x)^\top \Sigma a(x) + b(x)^2 \right\}$$

Let $\mathcal{M}(\mathbb{R})$ be the set of all the positive measures on \mathbb{R} . Then set \mathcal{P}_1 is equivalent to

$$\mathcal{P}_1 = \left\{ \mathbb{P} \in \mathcal{M}(\mathbb{R}) : \mathbb{E}_{\mathbb{P}}[1] = 1, \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] = b(x), \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}^2] = a(x)^\top \Sigma a(x) + b(x)^2 \right\}$$

where the first equality is to guarantee that \mathbb{P} is indeed a probability measure. Thus, the infimum in the set Z is equivalent to

$$\begin{aligned} & \inf_{\mathbb{P} \in \mathcal{M}(\mathbb{R})} \mathbb{E}[\mathbb{I}(|\boldsymbol{\xi}| \leq T)] \\ & \text{s.t. } \mathbb{E}_{\mathbb{P}}[1] = 1, \\ & \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] = b(x), \\ & \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}^2] = a(x)^\top \Sigma a(x) + b(x)^2, \end{aligned}$$

where $\mathbb{I}(\mathcal{R})$ is 1 if event \mathcal{R} is true, 0, otherwise. By dualizing three equality constraints above with dual multipliers λ, γ, β , Theorem 5.99 in [21] implies that for any $x \in Z$, $\inf_{\mathbb{P} \in \mathcal{P}_1} \mathbb{P}[|\boldsymbol{\xi}| \leq T] = \inf_{\mathbb{P} \in \mathcal{P}_1} \mathbb{E}[\mathbb{I}(|\boldsymbol{\xi}| \leq T)]$ can be reformulated as

$$\max_{\lambda, \gamma, \beta} \lambda + b(x)\gamma + (a(x)^\top \Sigma a(x) + b(x)^2)\beta \quad (6.11a)$$

$$\text{s.t. } \lambda + \xi\gamma + \xi^2\beta \leq 1, \forall \xi \in \mathbb{R} \quad (6.11b)$$

$$\lambda + \xi\gamma + \xi^2\beta \leq 0, \forall \xi : \xi > T \quad (6.11c)$$

$$\lambda + \xi\gamma + \xi^2\beta \leq 0, \forall \xi : \xi < -T. \quad (6.11d)$$

Note that in (6.11), we must have $\beta < 0$ and $|\frac{\gamma}{2\beta}| \leq T$. Otherwise, $\sup_{\xi} \lambda + \xi\gamma + \xi^2\beta \leq 0$ which implies that for any probability measure $\mathbb{P} \in \mathcal{P}_1$, we have

$$\lambda + b(x)\gamma + (a(x)^\top \Sigma a(x) + b(x)^2)\beta = \lambda + \mathbb{E}_{\mathbb{P}}[\gamma\boldsymbol{\xi} + \boldsymbol{\xi}^2\beta] \leq 0$$

contradiction that $x \in Z$.

Since $\beta < 0$ and $|\frac{\gamma}{2\beta}| \leq T$, (6.11) is equal to

$$\max_{\lambda, \gamma, \beta} \lambda + b(x)\gamma + (a(x)^\top \Sigma a(x) + b(x)^2)\beta \quad (6.12a)$$

$$\text{s.t. } \lambda - \frac{\gamma^2}{4\beta} \leq 1, \quad (6.12b)$$

$$\lambda + T|\gamma| + T^2\beta \leq 0, \quad (6.12c)$$

$$\beta < 0. \quad (6.12d)$$

Note that in (6.12), the optimal γ must have the same sign as $b(x)$ so as to maximize the objective function. Thus, set Z is equivalent to

$$Z = \left\{ x : \begin{cases} \lambda + |b(x)||\gamma| + b(x)^2\beta + a(x)^\top \Sigma a(x)\beta \geq 1 - \epsilon, \\ \lambda - \frac{\gamma^2}{4\beta} \leq 1, \\ \lambda + T|\gamma| + T^2\beta \leq 0, \\ \beta < 0. \end{cases} \right\} \quad (6.13)$$

Now in (6.13), let $\pi := -\frac{1}{\beta}$, $\hat{\gamma} := -|\gamma|/\beta$, $\hat{\lambda} := -\lambda/\beta$ and define a new set below

$$\bar{Z} = \left\{ x : \begin{cases} \hat{\lambda} + |b(x)|\hat{\gamma} - b(x)^2 - a(x)^\top \Sigma a(x) \geq (1 - \epsilon)\pi, \end{cases} \right. \quad (6.14a)$$

$$\left\{ x : \begin{cases} \hat{\lambda} + \frac{\hat{\gamma}^2}{4} \leq \pi, \end{cases} \right. \quad (6.14b)$$

$$\left\{ x : \begin{cases} \hat{\lambda} + T\hat{\gamma} - T^2 \leq 0, \end{cases} \right. \quad (6.14c)$$

$$\left\{ x : \begin{cases} \hat{\gamma} \geq 0. \end{cases} \right. \quad (6.14d)$$

Now we claim that $\bar{Z} = Z$. To prove this claim, we first show that $\bar{Z} \subseteq Z$. Given $x \in \bar{Z}$, there exists $(\hat{\lambda}, \hat{\gamma}, \pi)$ such that $(\hat{\lambda}, \hat{\gamma}, \pi, x)$ satisfy (6.14). First of all, by letting (6.14b) minus (6.14a), we have

$$\left(\frac{\hat{\gamma}}{2} - |b(x)| \right)^2 + a(x)^\top \Sigma a(x) \leq \epsilon\pi$$

thus, $\pi \geq 0$. There are two case:

Case 1. If $\pi = 0$, then by (6.14b) and (6.14a), we have $\hat{\lambda} = \hat{\gamma} = 0$ and $a(x)^\top \Sigma a(x) = 0, b(x)^2 = 0$. Next, choose $\lambda = 1, \gamma = 0, \beta = -T^{-2}$ and we have $(\lambda, \gamma, \beta, x)$ satisfies constraints in (6.13). Hence, $x \in Z$.

Case 2. If $\pi > 0$, now define $(\bar{\lambda}, \bar{\gamma}) = (\hat{\lambda}, \hat{\gamma})/\pi$ and $\beta = -1/\pi < 0$. Clearly, $(\bar{\lambda}, \bar{\gamma}, \beta, x)$ satisfy (6.13). Thus, $x \in Z$.

Thus $\bar{Z} \subseteq Z$. Next we show that $\bar{Z} \supseteq Z$. Given $x \in Z$, there exists (λ, γ, β) such that $(\lambda, \gamma, \beta, x)$ satisfy (6.13). As $\beta < 0$, let $(\hat{\lambda}, \hat{\gamma}) = -(\lambda, \gamma)/\beta$ and $\pi = -1/\beta$, then $(\hat{\lambda}, \hat{\gamma}, \pi, x)$ satisfy (6.14). Hence, $x \in \bar{Z}$. Thus $\bar{Z} = Z$.

By eliminating variables λ, π by Fourier-Motzkin procedure, (6.14) yields

$$Z = \left\{ x : \begin{array}{l} (|b(x)| - \gamma/2)^2 + a(x)^\top \Sigma a(x) \leq \epsilon(T - \gamma/2)^2, \\ \gamma \geq 0. \end{array} \right\} \quad (6.15a)$$

$$(6.15b)$$

Optimizing γ in (6.15) by distinguishing the cases $\epsilon T \geq |b(x)|$ ($\gamma^* = 0$), $b(x) \leq -\epsilon T$ ($\gamma^* = \frac{2}{1-\epsilon}(|b(x)| - \epsilon T)$) and $b(x) \geq \epsilon T$ ($\gamma^* = \frac{2}{1-\epsilon}(|b(x)| - \epsilon T)$), set Z can be reformulated as a disjunction of three sets Z_1, Z_2 and Z_3 . \square

We are now ready to prove Theorem 29. From the proof of Lemma 11, we observe that the best γ in (6.15) must be no larger than $2|b(x)|$; otherwise, we will arrive at a smaller set. Thus, (6.15) is equivalent to

$$Z = \left\{ x : \begin{array}{l} (|b(x)| - \pi)^2 + a(x)^\top \Sigma a(x) \leq \epsilon(T - \pi)^2, \\ 0 \leq \pi \leq |b(x)| \end{array} \right\} \quad (6.16a)$$

$$(6.16b)$$

where $\pi := \frac{\gamma}{2}$.

Now let \tilde{Z} denote the right-hand side in (6.7). We claim. $\tilde{Z} = Z$. Given $x \in Z$, there exists a π such that (π, x) satisfy (6.16). Now let $y = |b(x)| - \pi$, then (y, π, x) satisfy (6.7). Hence, $x \in \tilde{Z}$. Thus $\tilde{Z} \supseteq Z$. Next we show that $\tilde{Z} \subseteq Z$. Given $x \in \tilde{Z}$, there exists (y, π) such that (y, π, x) satisfy (6.7). There are two cases:

Case 1. if $|b(x)| \leq \pi \leq T$, then by (6.7), we have

$$a(x)^\top \Sigma a(x) \leq y^2 + a(x)^\top \Sigma a(x) \leq \epsilon(T - \pi)^2 \leq \epsilon(T - |b(x)|)^2 \quad (6.17)$$

where the first inequality is due to $y^2 \geq 0$, the second inequality is because of (6.7a) and the third inequality is because of $|b(x)| \leq \pi \leq T$. Now we distinguish whether $|b(x)| \leq \epsilon T$ or not.

a) if $|b(x)| \leq \epsilon T$, then by (6.7), we have

$$a(x)^\top \Sigma a(x) \leq \epsilon(T - \pi)^2 \leq \epsilon(T - |b(x)|)^2 \leq \epsilon T^2 - (2 - \epsilon)b(x)^2 \leq \epsilon T^2 - b(x)^2$$

where the first inequality is due to (6.7a), the second inequality is due to $|b(x)| \leq \pi \leq T$, the third inequality is due to $|b(x)| \leq \epsilon T$, and the last one is because of $\epsilon \in (0, 1)$. This leads to $b(x)^2 + a(x)^\top \Sigma a(x) \leq \epsilon T^2$, i.e., $x \in Z_1 \subseteq Z$.

b) if $|b(x)| \geq \epsilon T$, then (6.17) implies that

$$|b(x)| + \sqrt{\frac{1}{\epsilon}} \sqrt{a(x)^\top \Sigma a(x)} \leq T.$$

Since $\sqrt{\frac{1-\epsilon}{\epsilon}} \leq \sqrt{\frac{1}{\epsilon}}$, we must have $x \in Z_2 \cup Z_3 \subseteq Z$.

Case 2. if $0 \leq \pi \leq |b(x)|$, then $y \geq |b(x)| - \pi \geq 0$. Hence, (6.7) implies that

$$(|b(x)| - \pi)^2 + a(x)^\top \Sigma a(x) \leq y^2 + a(x)^\top \Sigma a(x) \leq \epsilon(T - \pi)^2$$

where the first inequality is due to $y \geq |b(x)| - \pi \geq 0$, second inequality is because of (6.7a) and thus (π, x) satisfies (6.16), i.e., $x \in Z$.

This completes the proof. \square

Using the above result we can now provide an exact SOCP formulation of DRCC-OPF (6.1) as follows:

$$v^* = \min_{\bar{p}, \alpha, \bar{\theta}} \sum_{i \in \mathcal{G}} \left(c_i (\bar{p}_i + r_i)^2 + c_i \alpha_i^2 e^\top \Sigma e \right) \quad (6.18a)$$

s.t. (6.1b) – (6.1d), (6.1g)

$$\begin{aligned} & \hat{y}_{ij}^2 + \beta_{ij}^2 \left(\check{\mathcal{B}}_{i\cdot}^\mathcal{W} - \alpha_i e - \check{\mathcal{B}}_{j\cdot}^\mathcal{W} + \alpha_j e \right)^\top \Sigma \left(\check{\mathcal{B}}_{i\cdot}^\mathcal{W} - \alpha_i e - \check{\mathcal{B}}_{j\cdot}^\mathcal{W} + \alpha_j e \right) \\ & \leq \hat{\epsilon}_{ij} (f_{ij}^{\max} - \hat{\pi}_{ij})^2, \forall (i, j) \in \mathcal{E}, \end{aligned} \quad (6.18b)$$

$$\beta_{ij} (\bar{\theta}_i - \bar{\theta}_j) \leq \hat{y}_{ij} + \hat{\pi}_{ij}, \forall (i, j) \in \mathcal{E}, \quad (6.18c)$$

$$\beta_{ij} (\bar{\theta}_j - \bar{\theta}_i) \leq \hat{y}_{ij} + \hat{\pi}_{ij}, \forall (i, j) \in \mathcal{E}, \quad (6.18d)$$

$$\hat{y}_{ij} \geq 0, 0 \leq \hat{\pi}_{ij} \leq f_{ij}^{\max}, \forall (i, j) \in \mathcal{E}, \quad (6.18e)$$

$$\bar{y}_i^2 + \alpha_i^2 e^\top \Sigma e \leq \bar{\epsilon}_i \left(\frac{p_i^{\max} - p_i^{\min}}{2} - \bar{\pi}_i \right)^2, \forall i \in \mathcal{G}, \quad (6.18f)$$

$$\bar{p}_i - \frac{p_i^{\max} + p_i^{\min}}{2} \leq \bar{y}_i + \bar{\pi}_i, \forall i \in \mathcal{G}, \quad (6.18g)$$

$$-\bar{p}_i + \frac{p_i^{\max} + p_i^{\min}}{2} \leq \bar{y}_i + \bar{\pi}_i, \forall i \in \mathcal{G}, \quad (6.18h)$$

$$0 \leq \bar{y}_i, 0 \leq \bar{\pi}_i \leq \frac{p_i^{\max} - p_i^{\min}}{2}, \forall i \in \mathcal{G}, \quad (6.18i)$$

where $\hat{\pi}, \hat{y}, \bar{\pi}, \bar{y}$ are auxiliary nonnegative variables.

6.4.3 Quality of approximation of Z by $Z_A(\epsilon/2), Z_A(\epsilon)$

We know that from Theorem 28, we have $Z_A(\epsilon/2) \subseteq Z \subseteq Z_A(\epsilon)$ and usually the inclusion is strict. The following example shows that the distances between set Z and $Z_A(\epsilon)$ and between set Z and $Z_A(\epsilon/2)$ can be large.

Example 9. Let $b(x) = 0, \Sigma = I$, then

$$\begin{aligned} Z_A(\epsilon/2) &= \left\{ x : \|a(x)\|_2^2 \leq \frac{\epsilon}{2-\epsilon} T^2 \right\} \\ Z_A(\epsilon) &= \left\{ x : \|a(x)\|_2^2 \leq \frac{\epsilon}{1-\epsilon} T^2 \right\} \\ Z &= \{x : \|a(x)\|_2^2 \leq \epsilon T^2\} \end{aligned}$$

Clearly, when $\epsilon \rightarrow 1$, $Z_A(\epsilon) \rightarrow \mathbb{R}^n$ but Z is close to a ball $\{x : \|a(x)\|_2^2 \leq T^2\}$. Hence the distance between Z and $Z_A(\epsilon)$ tends to be infinity.

On the other hand, we know that $\frac{\epsilon}{2-\epsilon} \approx \frac{\epsilon}{2}$ when ϵ is small. Thus, in this case, the radius of ball Z could be almost $\sqrt{2}$ larger than $Z_A(\epsilon/2)$. This inner approximation could easily lead the feasible region of a DRCC-OPF to be infeasible. For example, if there is an additional constraint $S = \{x : a(x) \geq T\sqrt{\frac{2\epsilon}{3m}}e\}$ where m is the dimension of $a(x)$, then clearly $S \cap Z_A(\epsilon/2) = \emptyset$ when $\epsilon < 0.5$, but $S \cap Z$ even has a nonempty interior.

6.5 Numerical Illustration

We test the DRCC-OPF model (6.18) with an example used in [19]: case39 of MATPOWER data originally from [134]. The case is available at <http://www.pserc.cornell.edu/matpower/>. In this data set, there are 39 buses (set \mathcal{V}), 46 lines (set \mathcal{E}) and 10 generators (set \mathcal{G}). We assume that renewable power can be generated from buses 1 to 4 (set \mathcal{W}) with mean $\mu_i = 40$ (MW) for each $i \in \mathcal{W}$ and its covariance matrix Σ is diagonal with $\Sigma(i, i) = 400$ for each $i \in \mathcal{W}$. All of the instances are solved by CVX [43].

In our first test, we let $\hat{\epsilon}_{ij} = \bar{\epsilon}_i = 0.2$. We compare our method with a “risk neutral” model by assuming there is no uncertainty in (6.1e) and (6.1f), i.e., reformulating these

constraints as

$$|\beta_{ij}(\bar{\theta}_i - \bar{\theta}_j)| \leq f_{ij}^{\max}, \forall (i, j) \in \mathcal{E}, \quad (1e')$$

$$p_i^{\min} \leq \bar{p}_i \leq p_i^{\max}, \forall i \in \mathcal{G}; \quad (1f')$$

and to the model in [19] (we call it “*BCH model*”) where they assume the underlying distribution is Gaussian. All three models can be solved within a second, with total costs 36059.1, 36448.6, 37885.3 for the risk neutral model, BCH model, and our model (6.18), respectively. Thus, there is no significant difference (within 5%) of total costs among all the three models.

We also test the reliability of models by simulating different distributions of renewables’ output, i.e., Gaussian, student, Laplace, Logistic and uniform distributions. We generate 100,000 samples from each distribution and check the violation of line flow capacity and bus capacity for each transmission line and bus. In Table 9, we compute the maximum probability of violations across all the lines and buses under each distribution. It can be seen that even when the risk parameters are all equal to 0.2, our model is quite robust and the chance that a line or bus capacity will be violated is close to zero for most of distributions. However, in the risk neutral model, there is a 50% chance that a line or bus capacity is violated almost for each distribution. The BCH model does slightly better, but still under some distributions (e.g., Logistic), the probability of failure is relatively high (31%).

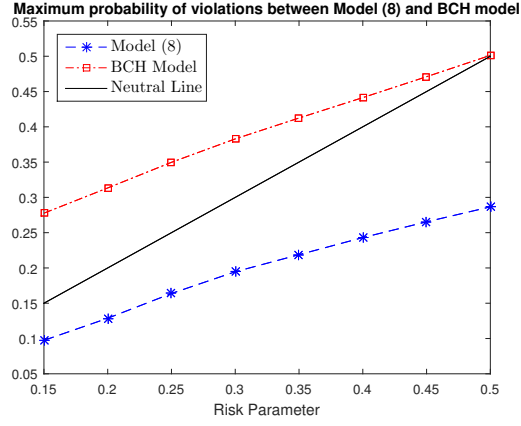
Table 9: Maximum probability of violations and total costs among model (6.18), risk neutral model and BCH model

Distribution	Model (6.18)	Risk Neutral	BCH
Total Cost	37885.3	36059.1	36448.6
Gaussian	0.02279	0.50149	0.2022
Student	1.00E-05	0.50128	5.00E-05
Laplace	0.0274	0.50366	0.18197
Logisitic	0.12856	0.5022	0.31184
Uniform	0.0211	0.5016	0.21614

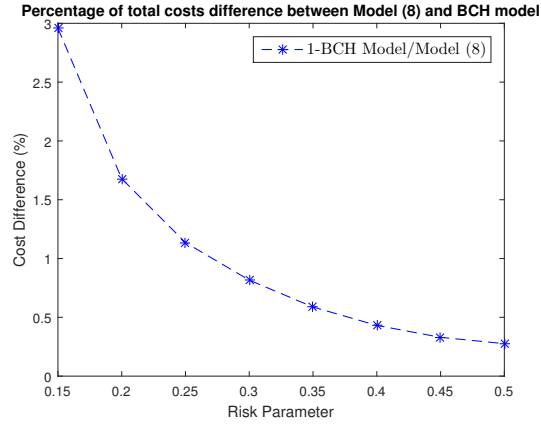
In the second test, we let the risk parameters $\hat{\epsilon}_{ij}$ and $\bar{\epsilon}_i$ range from 0.15 to 0.5 and observe how this affects the solutions. We compare our results on maximum probability of

violations with the ones of BCH model through generating 100,000 samples from Logistic distribution. In Figure 6(a), we see that the results from BCH model are quite sensitive to the risk parameters, i.e., the probability of violating line capacity or bus capacity increases almost linearly as the risk parameters grows. Since the probability of violation curve is always above the neutral line which tells whether the probability of violations is larger than the prespecified risk parameter ϵ or not, hence the solution of BCH is not robust at all. Therefore, in the BCH model, one might need to stick to small risk parameters. Our model (6.18) turns out to be quite robust with the risk parameters. Even when all of the risk parameters are equal to 0.5, the chance of capacity violation is still quite small (around 28%). We also observe that in Figure 6(b), cost difference between two models reduces when the risk parameter increases. Another observation is that the total cost of our model (6.18) is the most costly due to its conservativeness, but the difference between ours and risk-free model is small (at most 6%) . This could be because in the objective function (6.1a), there is only production cost of regular generators but no cost on renewables. Hence influence of renewables to the total costs is small but to the system reliability is dramatic. On the other hand, if the operators would like to reduce the total costs and can tolerate a relatively high risk, they can increase the risk parameter.

Finally, we compare the computational time for model (6.18) with that of the risk neutral model and the BCH model, by solving different MATPOWER cases: case30, case39, case57, case118 and case145. The sizes of these instances and the associated run times are shown in Table 10. The results in Table 10 show that even for the large-size power network (case145), all three models can be solved efficiently (i.e., within 4 seconds). We also observe that even though model (6.18) requires more variables – at most $2\mathcal{E} + 2\mathcal{G}$ additional variables – its computational time is similar to the other two models.



(a) Max. Prob. of violations



(b) Total costs

Figure 6: Comparison between model (6.18) and BCH model

Table 10: Computational time comparison among model (6.18), risk neutral model and BCH model

Cases		case30	case39	case57	case118	case145
Data	Buses ($ \mathcal{V} $)	30	39	57	118	145
	Lines ($ \mathcal{E} $)	41	46	80	186	453
	Generators ($ \mathcal{G} $)	6	10	7	54	50
	Renewables ($ \mathcal{W} $)	3	3	5	11	14
Results	Model (6.18)	1.48	1.13	1.02	2.08	3.35
	Risk Neutral	0.83	0.57	1.36	1.71	1.63
	BCH	0.83	0.92	1.63	1.60	3.41

CHAPTER VII

OPTIMIZED BONFERRONI APPROXIMATIONS OF DISTRIBUTIONALLY ROBUST JOINT CHANCE CONSTRAINTS

7.1 Introduction

7.1.1 Setting

A linear chance constrained optimization problem is of the form:

$$\min \quad c^\top x, \tag{7.1a}$$

$$\text{s.t.} \quad x \in S, \tag{7.1b}$$

$$\mathbb{P} \left\{ \boldsymbol{\xi} : a_i(x)^\top \boldsymbol{\xi}_i \leq b_i(x), \forall i \in [I] \right\} \geq 1 - \epsilon. \tag{7.1c}$$

Above, the vector $x \in \mathbb{R}^n$ denotes the decision variables; the vector $c \in \mathbb{R}^n$ denotes the objective function coefficients; the set $S \subseteq \mathbb{R}^n$ denotes deterministic constraints on x ; and the constraint (7.1c) is a chance constraint involving I inequalities with uncertain data specified by the random vector $\boldsymbol{\xi}$ supported on a closed convex set $\Xi \subseteq \mathbb{R}^m$ with a known probability distribution \mathbb{P} . We let $[R] := \{1, 2, \dots, R\}$ for any positive integer R , and for each uncertain constraint $i \in [I]$, $a_i(x) \in \mathbb{R}^{m_i}$ and $b_i(x) \in \mathbb{R}$ denote affine mappings of x such that $a_i(x) = A^i x + a^i$ and $b_i(x) = B^i x + b^i$ with $A^i \in \mathbb{R}^{m_i \times n}$, $a^i \in \mathbb{R}^{m_i}$, $B^i \in \mathbb{R}^n$, and $b^i \in \mathbb{R}$, respectively. The uncertain data associated with constraint i is specified by $\boldsymbol{\xi}_i$ which is the projection of $\boldsymbol{\xi}$ to a coordinate subspace $\mathcal{S}_i \subseteq \mathbb{R}^m$, i.e., \mathcal{S}_i is a span of m_i distinct standard bases with $\dim(\mathcal{S}_i) = m_i$. The support of $\boldsymbol{\xi}_i$ is $\Xi_i = \text{Proj}_{\mathcal{S}_i}(\Xi)$. The chance constraint (7.1c) requires that all I uncertain constraints are simultaneously satisfied with a probability or reliability level of at least $(1 - \epsilon)$, where $\epsilon \in (0, 1)$ is a specified risk tolerance. We call (7.1c) a *single* chance constraint if $I = 1$ and a *joint* chance constraint if $I \geq 2$.

Remark 4. *The notation above might appear to indicate that the uncertain data is separable across the inequalities. However, note that $\boldsymbol{\xi}_i$ is a projection of $\boldsymbol{\xi}$. For example, we may*

have $\xi_i = \xi$ and $\mathcal{S}_i = \mathbb{R}^m$ for all i , when each inequality involves all uncertain coefficients ξ .

In practice, the decision makers often have limited distributional information on ξ , making it challenging to commit to a single \mathbb{P} . As a consequence, the optimal solution to (7.1a)–(7.1c) can actually perform poorly if the (true) probability distribution of ξ is different from the one we commit to in (7.1c). In this case, a natural alternative of (7.1c) is a distributionally robust chance constraint of the form

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left\{ \xi : a_i(x)^\top \xi_i \leq b_i(x), \forall i \in [I] \right\} \geq 1 - \epsilon, \quad (7.1d)$$

where we specify a family \mathcal{P} of probability distributions of ξ , called an *ambiguity set*, and the chance constraint (7.1c) is required to hold for all the probability distributions \mathbb{P} in \mathcal{P} . We call formulation (7.1a)–(7.1b), (7.1d) a distributionally robust joint chance constrained program (DRCCP) and denote the feasible region induced by (7.1d) as

$$Z := \left\{ x \in \mathbb{R}^n : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left\{ \xi : a_i(x)^\top \xi_i \leq b_i(x), \forall i \in [I] \right\} \geq 1 - \epsilon \right\}. \quad (7.2)$$

In general, the set Z is nonconvex and leads to NP-hard optimization problems [46]. This is not surprising since the same conclusion holds even when the ambiguity set \mathcal{P} is a singleton [68, 73]. The focus of this chapter is on developing tractable convex approximations and reformulations of set Z .

7.1.2 Related Literature

Existing literature has identified a number of important special cases where Z is convex. In the non-robust setting, i.e. when \mathcal{P} is a singleton, the set Z is convex if $A^i = 0$ for all $i \in [I]$ (i.e. the uncertainties do not affect the variable coefficients) and either (i) the distribution of the vector $[(a^1)^\top \xi_1, \dots, (a^I)^\top \xi_I]^\top$ is quasi-concave [81, 114, 115] or (ii) the components of vector $[(a^1)^\top \xi_1, \dots, (a^I)^\top \xi_I]^\top$ are independent and follow log-concave probability distributions [80]. Much less is known about the case $A^i \neq 0$ (i.e. with uncertain coefficients), except that Z is convex if $I = 1$, $\epsilon \leq 1/2$, and ξ has a symmetric and non-degenerate log-concave distribution [57], of which the normal distribution is a special case

[53]. In the robust setting, when \mathcal{P} consists of all probability distributions with given first and second moments and $I = 1$, the set Z is second-order cone representable [24, 37]. Similar convexity results hold when $I = 1$ and \mathcal{P} also incorporates other distributional information such as the support of ξ [32], the unimodality of \mathbb{P} [46, 59], or arbitrary convex mapping of ξ [118]. For distributionally robust *joint* chance constraints, i.e. $I \geq 2$ and \mathcal{P} is not a singleton, conditions for convexity of Z are scarce. To the best of our knowledge, [47] provides the first convex reformulation of Z in the absence of coefficient uncertainty, i.e. $A^i = 0$ for all $i \in [I]$, when \mathcal{P} is characterized by the mean, a positively homogeneous dispersion measure, and a conic support of ξ . For the more general coefficient uncertainty setting, i.e. $A^i \neq 0$, [118] identifies several sufficient conditions for Z to be convex (e.g., when \mathcal{P} is specified by one moment constraint), and [117] shows that Z is convex when the chance constraint (7.1d) is two-sided (i.e., when $I = 2$ and $a_1(x)^\top \xi_1 = -a_2(x)^\top \xi_2$) and \mathcal{P} is characterized by the first two moments.

Various approximations have been proposed for settings where Z is not convex. When \mathcal{P} is a singleton, i.e. $\mathcal{P} = \{\mathbb{P}\}$, [73] propose a family of deterministic convex inner approximations, among which the conditional-value-at-risk (CVaR) approximation [86] is proved to be the tightest. A similar approach is used to construct convex outer approximations in [3]. Sampling based approaches that approximate the true distribution by an empirical distribution are proposed in [23, 67, 74]. When the probability distribution \mathbb{P} is discrete, [4] develop Lagrangian relaxation schemes and corresponding primal linear programming formulations. In the distributionally robust setting, [30] propose to aggregate the multiple uncertain constraints with positive scalars in to a single constraint, and then use CVaR to develop an inner approximation of Z . This approximation is shown to be exact for distributionally robust single chance constraints when \mathcal{P} is specified by first and second moments in [135] or, more generally, by convex moment constraints in [118].

7.1.3 Contributions

In this chapter we study the set Z in the distributionally robust joint chance constraint setting, i.e. $I \geq 2$ and \mathcal{P} is not a singleton. In particular, we consider a classical approximation scheme for joint chance constraint, termed Bonferroni approximation [30, 73, 135]. This scheme decomposes the joint chance constraint (7.1d) into I single chance constraints where the risk tolerance of constraint i is set to a fixed parameter $s_i \in [0, \epsilon]$ such that $\sum_{i \in [I]} s_i \leq \epsilon$. Then, by the union bound, it is easy to see that any solution satisfying all I single chance constraints will satisfy the joint chance constraint. Such a distributionally robust single chance constraint system is often significantly easier than the joint constraint. To optimize the quality of the Bonferroni approximation, it is attractive to treat $\{s_i\}_{i \in [I]}$ as design variables rather than as fixed parameters. However, this could undermine the convexity of the resulting approximate system and make it challenging to solve. Indeed, [73] cites the tractability of this *optimized* Bonferroni approximation as “an open question” (see Remark 2.1 in [73]). In this chapter, we make the following contributions to the study of optimized Bonferroni approximation:

1. We show that the optimized Bonferroni approximation of a distributionally robust joint chance constraint is in fact *exact* when the uncertainties are separable across the individual inequalities, i.e., each uncertain constraint involves a different set of uncertain parameters and corresponding distribution families.
2. For the setting when the ambiguity set is specified by the first two moments of the uncertainties in each constraint, we establish that the optimized Bonferroni approximation, in general, leads to strongly NP-hard problems; and go on to identify several sufficient conditions under which it becomes tractable.
3. For the setting when the ambiguity set is specified by marginal distributions of the uncertainties in each constraint, again, we show that while the general case is strongly NP-hard, there are several sufficient conditions leading to tractability.
4. For moment based distribution families and binary decision variables, we show that the

optimized Bonferroni approximation can be reformulated as a mixed integer second-order conic set.

5. Finally, we demonstrate how our results can be used to derive a convex reformulation of a distributionally robust joint chance constraint with a specific non-separable distribution family.

7.2 Optimized Bonferroni Approximation

In this section we formally present the optimized Bonferroni approximation of the distributionally robust joint constraint set Z , compare it with alternative single chance constraint approximations, and provide a sufficient condition under which it is exact.

7.2.1 Single chance constraint approximations

Recall that the uncertain data associated with constraint $i \in [I]$ is specified by ξ_i which is the projection of ξ to a coordinate subspace $\mathcal{S}_i \subseteq \mathbb{R}^m$ with $\dim(\mathcal{S}_i) = m_i$, and the support of ξ_i is $\Xi_i = \text{Proj}_{\mathcal{S}_i}(\Xi)$. For each $i \in [I]$, let \mathcal{D}_i denote the projection of the ambiguity set \mathcal{P} to the coordinate subspace \mathcal{S}_i , i.e., $\mathcal{D}_i = \text{Proj}_{\mathcal{S}_i}(\mathcal{P})$. Thus \mathcal{D}_i denotes the projected ambiguity set associated with the uncertainties appearing in constraint i . The following two examples illustrate ambiguity set \mathcal{P} and its projections $\{\mathcal{D}_i\}_{i \in [I]}$.

Example 10. *Consider*

$$Z = \left\{ x \in \mathbb{R}^2 : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left\{ \xi : \begin{array}{l} \hat{\xi}_1 x_1 + \hat{\xi}_2 x_2 \leq 0 \\ \hat{\xi}_2 x_1 + \hat{\xi}_3 x_2 \leq 1 \end{array} \right\} \geq 0.75 \right\},$$

where $\xi = [\hat{\xi}_1, \hat{\xi}_2, \hat{\xi}_3]^\top$, $\xi_1 = [\hat{\xi}_1, \hat{\xi}_2]^\top$, $\xi_2 = [\hat{\xi}_2, \hat{\xi}_3]^\top$, and $\mathcal{P} = \{\mathbb{P} : \mathbb{E}_{\mathbb{P}}[\xi] = 0, \mathbb{E}_{\mathbb{P}}[\xi\xi^\top] = \Sigma\}$ with

$$\Sigma = \begin{bmatrix} 1 & 0 & 1.2 \\ 0 & 0.5 & 0.5 \\ 1.2 & 0.5 & 2 \end{bmatrix}.$$

In this example, $m = 3$, $m_1 = m_2 = 2$, $\mathcal{S}_1 = \{\hat{\xi} \in \mathbb{R}^3 : \hat{\xi}_3 = 0\}$, $\mathcal{S}_2 = \{\hat{\xi} \in \mathbb{R}^3 : \hat{\xi}_1 = 0\}$, and

$\mathcal{D}_i = \{\mathbb{P} : \mathbb{E}_{\mathbb{P}}[\xi_i] = 0, \mathbb{E}_{\mathbb{P}}[\xi_i \xi_i^\top] = \Sigma_i\}$ for $i = 1, 2$, where

$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix} \quad \text{and} \quad \Sigma_2 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 2 \end{bmatrix}. \quad \square$$

Example 11. Consider

$$Z = \left\{ x \in \mathbb{R}^I : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \{ \xi : \xi_i \leq x_i, \forall i \in [I] \} \geq 0.9 \right\},$$

where $\xi \sim \mathcal{N}(\mu, \Sigma)$, i.e. \mathcal{P} is a singleton containing only an I -dimensional multivariate normal distribution with mean $\mu \in \mathbb{R}^I$ and covariance matrix $\Sigma \in \mathbb{R}^{I \times I}$. In this example, $m = I$, and for all $i \in [I]$, $m_i = 1$, $\mathcal{S}_i = \{\xi \in \mathbb{R}^I : \xi_j = 0, j \neq i, \forall j \in [I]\}$, and \mathcal{D}_i is a singleton containing only a 1-dimensional normal distribution with mean μ_i and variance Σ_{ii} . \square

Consider the following two distributionally robust single chance constraint approximations of Z :

$$Z_O := \left\{ x \in \mathbb{R}^n : \inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i \left\{ \xi_i : a_i(x)^\top \xi_i \leq b_i(x) \right\} \geq 1 - \epsilon, \forall i \in [I] \right\}, \quad (7.3)$$

and

$$Z_I := \left\{ x \in \mathbb{R}^n : \inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i \left\{ \xi_i : a_i(x)^\top \xi_i \leq b_i(x) \right\} \geq 1 - \frac{\epsilon}{I}, \forall i \in [I] \right\}. \quad (7.4)$$

Both Z_O and Z_I involve I distributionally robust single chance constraints, and they differ by the choice of the risk levels. The approximation Z_O relaxes the requirement of simultaneously satisfying all uncertain linear constraints, and hence is an outer approximation of Z . In Z_I , each single chance constraint has a risk level of ϵ/I , and it follows from the union bound (or Bonferroni inequality[20]), that Z_I is an inner approximation of Z . The set Z_I is typically called the Bonferroni approximation. We consider an extension of Z_I where the risk level of each constraint is not fixed but optimized [73]. The resulting optimized Bonferroni approximation is:

$$Z_B := \left\{ x : \inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i \left\{ \xi_i : a_i(x)^\top \xi_i \leq b_i(x) \right\} \geq 1 - s_i, s_i \geq 0, \forall i \in [I], \sum_{i \in [I]} s_i \leq \epsilon \right\}. \quad (7.5)$$

7.2.2 Comparison of Approximation Schemes

From the previous discussion we know that Z_O is an outer approximation of Z , while both Z_B and Z_I are inner approximations of Z and that Z_B is at least as tight as Z_I . We formalize this observation in the following result.

Theorem 30. $Z_O \supseteq Z \supseteq Z_B \supseteq Z_I$.

Proof. By construction, $Z_O \supseteq Z$. To show that $Z \supseteq Z_B$, we pick $x \in Z_B$. For all $\mathbb{P} \in \mathcal{P}$ and $i \in [I]$, $x \in Z_B$ implies that $\mathbb{P}\{\boldsymbol{\xi} : a_i(x)^\top \boldsymbol{\xi}_i \leq b_i(x)\} = \mathbb{P}_i\{\boldsymbol{\xi}_i : a_i(x)^\top \boldsymbol{\xi}_i \leq b_i(x)\} \geq 1 - s_i$, or equivalently, $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\{\boldsymbol{\xi} : a_i(x)^\top \boldsymbol{\xi}_i > b_i(x)\} \leq s_i$. Hence,

$$\begin{aligned} \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\{\boldsymbol{\xi} : a_i(x)^\top \boldsymbol{\xi}_i \leq b_i(x), \forall i \in [I]\} &= 1 - \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\{\boldsymbol{\xi} : \exists i \in [I], \text{ s.t. } a_i(x)^\top \boldsymbol{\xi}_i > b_i(x)\} \\ &\geq 1 - \sup_{\mathbb{P} \in \mathcal{P}} \sum_{i \in [I]} \mathbb{P}\{\boldsymbol{\xi} : a_i(x)^\top \boldsymbol{\xi}_i > b_i(x)\} \\ &\geq 1 - \sum_{i \in [I]} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\{\boldsymbol{\xi} : a_i(x)^\top \boldsymbol{\xi}_i > b_i(x)\} \\ &\geq 1 - \sum_{i \in [I]} s_i \geq 1 - \epsilon, \end{aligned}$$

where the first inequality is due to the Bonferroni inequality or union bound, the second inequality is because the supremum over summation is no larger than the sum of supremum, and the final inequality follows from the definition of Z_B . Thus, $x \in Z$. Finally, note that Z_I is a restriction of Z_B by setting $s_i = \epsilon/I$ for all $i \in [I]$ and so $Z_B \supseteq Z_I$. \square \square

The following example shows that all inclusions in Theorem 30 can be strict.

Example 12. Consider

$$Z = \left\{ x \in \mathbb{R}^2 : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left\{ \boldsymbol{\xi} : \begin{cases} \boldsymbol{\xi}_1 \leq x_1 \\ \boldsymbol{\xi}_2 \leq x_2 \\ x_1 \leq 1 \\ x_2 \leq 1 \end{cases} \right\} \geq 0.5 \right\},$$

where \mathcal{P} is a singleton containing the probability distribution that $\boldsymbol{\xi}^1$ and $\boldsymbol{\xi}^2$ are independent and uniformly distributed on $[0, 1]$. It follows that

$$Z_O = \{x \in [0, 1]^2 : x_1 \geq 0.5, x_2 \geq 0.5\},$$

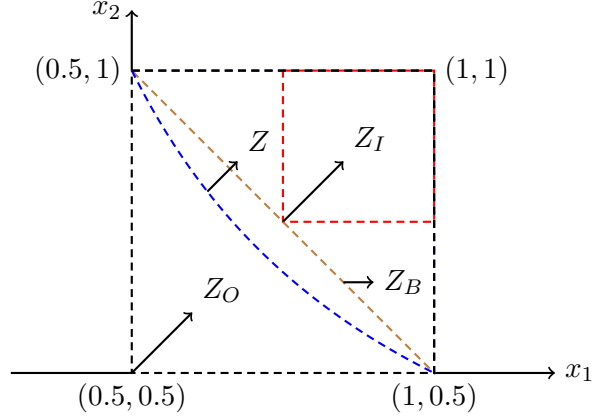


Figure 7: Illustration of Example 12

$$Z = \{x \in [0, 1]^2 : x_1 x_2 \geq 0.5\},$$

$$Z_B = \{x \in [0, 1]^2 : x_1 + x_2 \geq 1.5\}, \text{ and}$$

$$Z_I = \{x \in [0, 1]^2 : x_1 \geq 0.75, x_2 \geq 0.75\}.$$

We display these four sets in Fig. 7, where the dashed lines denotes the boundaries of Z_O, Z, Z_B, Z_I . It is clear that $Z_O \supsetneq Z \supsetneq Z_B \supsetneq Z_I$. \square

7.2.3 Exactness of Optimized Bonferroni Approximation

In this section we use a result from [89] to establish a sufficient condition under which the optimized Bonferroni approximation is exact. We first review this result.

Let $\{(\Xi_i, \mathcal{F}_i, \mathbb{P}_i) : i \in [I]\}$ be a finite collection of probability spaces, where for $i \in [I]$, $\Xi_i \subseteq \mathcal{S}_i$ is a sample space, \mathcal{F}_i is a σ -algebra of Ξ_i , and \mathbb{P}_i is a probability measure on (Ξ_i, \mathcal{F}_i) . Consider the product space $(\Xi, \mathcal{F}) = \prod_{i \in [I]} (\Xi_i, \mathcal{F}_i)$, and let $\mathcal{M}(\Xi, \mathcal{F})$ denote the set of all probability measures on (Ξ, \mathcal{F}) . Let $\mathcal{M}(\mathbb{P}_1, \dots, \mathbb{P}_I)$ denote the set of joint probability measures on (Ξ, \mathcal{F}) generated by $(\mathbb{P}_1, \dots, \mathbb{P}_I)$, i.e.

$$\mathcal{M}(\mathbb{P}_1, \dots, \mathbb{P}_I) = \{\mathbb{P} \in \mathcal{M}(\Xi, \mathcal{F}) : \text{Proj}_i(\mathbb{P}) = \mathbb{P}_i \ \forall i \in [I]\},$$

where $\text{Proj}_i : \Xi \rightarrow \Xi_i$ denotes the i -th projection operation. For any $\mathbb{P} \in \mathcal{M}(\mathbb{P}_1, \dots, \mathbb{P}_I)$, the Fréchet inequality [39], is:

$$\left[\sum_{i \in [I]} \mathbb{P}_i \{A_i\} - (I - 1) \right]_+ \leq \mathbb{P} \left\{ \prod_{i \in [I]} A_i \right\},$$

where $[a]_+ = \max\{0, a\}$.

Remark 5. Note that in the special case of $\Xi_i = \Xi$ for all $i \in [I]$, the above Fréchet inequality is

$$\left[\sum_{i \in [I]} \mathbb{P}_i \{A_i\} - (I - 1) \right]_+ \leq \mathbb{P} \left\{ \bigcap_{i \in [I]} A_i \right\},$$

which is essentially Bonferroni inequality complemented.

The following result establishes a tight version of the Fréchet inequality.

Theorem 31. (Theorem 6 in [89]) Let $\{(\Xi_i, \mathcal{F}_i) : i \in [I]\}$ be a finite collection of Polish spaces with associated probability measures $\{\mathbb{P}_1, \dots, \mathbb{P}_I\}$. Then for all $A_i \in \Xi_i$ with $i \in [I]$ it holds that

$$\left[\sum_{i \in [I]} \mathbb{P}_i \{A_i\} - (I - 1) \right]_+ = \inf \left\{ \mathbb{P} \left\{ \prod_{i \in [I]} A_i \right\} : \mathbb{P} \in \mathcal{M}(\mathbb{P}_1, \dots, \mathbb{P}_I) \right\}.$$

Next we use the above result to show that the optimized Bonferroni approximation Z_B , consisting single chance constraints, is identical to Z consisting of a joint chance constraint when the uncertainties in each constraint are *separable*, i.e. each uncertain constraint involves a different set of uncertain parameters and associated ambiguity sets. Recall that uncertain data in Z is described the random vector $\boldsymbol{\xi}$ supported on a closed convex set $\Xi \subseteq \mathbb{R}^m$, and the uncertain data associated with constraint i is specified by $\boldsymbol{\xi}_i$ which is the projection of $\boldsymbol{\xi}$ to a coordinate subspace $\mathcal{S}_i \subseteq \mathbb{R}^m$ with $\dim(\mathcal{S}_i) = m_i$. The support of $\boldsymbol{\xi}_i$ is $\Xi_i = \text{Proj}_{\mathcal{S}_i}(\Xi)$. Furthermore, the ambiguity set associated with the uncertainties appearing in constraint i , \mathcal{D}_i , is the projection of the ambiguity set \mathcal{P} to the coordinate subspace \mathcal{S}_i , i.e., $\mathcal{D}_i = \text{Proj}_{\mathcal{S}_i}(\mathcal{P})$. The separable uncertainty condition can then be formalized as follows:

(A1) $\Xi = \prod_{i \in [I]} \Xi_i$ and $\mathcal{P} = \prod_{i \in [I]} \mathcal{D}_i$, i.e., $\mathbb{P} \in \mathcal{P}$ if and only if $\text{Proj}_i(\mathbb{P}) \in \mathcal{D}_i$ for all $i \in [I]$.

The following example illustrates Assumption (A1).

Example 13. Consider

$$Z = \left\{ x \in \mathbb{R}^2 : \inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left\{ \boldsymbol{\xi} : \begin{array}{l} \boldsymbol{\xi}_1 \leq x_1 \\ 2\boldsymbol{\xi}_2 \leq x_1 + x_2 \end{array} \right\} \geq 0.75 \right\},$$

where $\Xi_1 = \mathbb{R}, \Xi_2 = \mathbb{R}, \Xi = \mathbb{R}^2$ and

$$\mathcal{P} = \{\mathbb{P} : \mathbb{E}_{\mathbb{P}}[\xi_1] = 0, \mathbb{E}_{\mathbb{P}}[\xi_1^2] = \sigma_1^2, \mathbb{E}_{\mathbb{P}}[\xi_2] = 0, \mathbb{E}_{\mathbb{P}}[\xi_2^2] = \sigma_2^2\}$$

$$\mathcal{D}_1 = \{\mathbb{P}_1 : \mathbb{E}_{\mathbb{P}_1}[\xi_1] = 0, \mathbb{E}_{\mathbb{P}_1}[\xi_1^2] = \sigma_1^2\}$$

$$\mathcal{D}_2 = \{\mathbb{P}_2 : \mathbb{E}_{\mathbb{P}_2}[\xi_2] = 0, \mathbb{E}_{\mathbb{P}_2}[\xi_2^2] = \sigma_2^2\}.$$

Clearly, $\Xi = \Xi_1 \times \Xi_2$ and $\mathcal{P} = \mathcal{D}_1 \times \mathcal{D}_2$. □

We are now ready to establish the exactness of optimized Bonferroni approximation under the above condition.

Theorem 32. *Under Assumption (A1), $Z = Z_B$.*

Proof. We have $Z_B \subseteq Z$ by Theorem 30. It remains to show that $Z \subseteq Z_B$. Given an $x \in Z$, we rewrite the left-hand side of (7.1d) as

$$\inf_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left\{ \xi : a_i(x)^\top \xi_i \leq b_i(x), \forall i \in [I] \right\} \quad (7.6a)$$

$$= \inf_{\mathbb{P}_i \in \mathcal{D}_i, \forall i \in [I]} \inf_{\mathbb{P} \in \mathcal{M}(\mathbb{P}_1, \dots, \mathbb{P}_I)} \mathbb{P} \left\{ \xi : a_i(x)^\top \xi_i \leq b_i(x), \forall i \in [I] \right\} \quad (7.6b)$$

$$= \inf_{\mathbb{P}_i \in \mathcal{D}_i, \forall i \in [I]} \left[\sum_{i \in [I]} \mathbb{P}_i \left\{ \xi_i : a_i(x)^\top \xi_i \leq b_i(x) \right\} - (I - 1) \right]_+, \quad (7.6c)$$

where equality (7.6b) decomposes the optimization problem in (7.6a) into two layers: the outer layer searches for optimal (i.e., worst-case) marginal distributions $\mathbb{P}_i \in \mathcal{D}_i$ for all $i \in [I]$, while the inner layer searches for the worst-case joint probability distribution that admits the given marginals \mathbb{P}_i . Equality (7.6c) follows from Theorem 31. Note that our sample space is Euclidean and is hence a Polish space. Since $x \in Z$, the right-hand-side of (7.6c) is no smaller than $1 - \epsilon$. It follows that (7.6c) is equivalent to

$$\begin{aligned} & \inf_{\mathbb{P}_i \in \mathcal{D}_i, \forall i \in [I]} \left[\sum_{i \in [I]} \mathbb{P}_i \left\{ \xi_i : a_i(x)^\top \xi_i \leq b_i(x) \right\} - (I - 1) \right] \\ &= \sum_{i \in [I]} \inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i \left\{ \xi_i : a_i(x)^\top \xi_i \leq b_i(x) \right\} - (I - 1), \end{aligned} \quad (7.6d)$$

where equality (7.6d) is because the ambiguity sets \mathcal{D}_i , $i \in [I]$, are separable by Assumption (A1). Finally, let $s_i := 1 - \inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i \left\{ \xi_i : a_i(x)^\top \xi_i \leq b_i(x) \right\}$ and so $s_i \geq 0$ for all $i \in [I]$.

Since $x \in Z$, by (7.6d), we have

$$\sum_{i \in [I]} (1 - s_i) - (I - 1) \geq 1 - \epsilon$$

which implies $\sum_{i \in [I]} s_i \leq \epsilon$. Therefore, $x \in Z_B$. \square \square

The above result establishes that if the ambiguity set of a distributionally robust joint chance constraint is specified in a form that is separable over the uncertain constraints, then the optimized Bonferroni approximation involving a system of distributionally robust single chance constraints is exact. In the next two sections, we investigate two such settings.

7.3 Ambiguity Set Based on the First Two Moments

In this section, we study the computational tractability of optimized Bonferroni approximation when the ambiguity set is specified by the first two moments of the projected random vectors $\{\xi_i\}_{i \in [I]}$. More specifically, for each $i \in [I]$, we make the following assumption on \mathcal{D}_i , the projection of the ambiguity set \mathcal{P} to the coordinate subspace \mathcal{S}_i :

(A2) The projected ambiguity sets $\{\mathcal{D}_i\}_{i \in [I]}$ are defined by the first and second moments of ξ_i :

$$\mathcal{D}_i = \left\{ \mathbb{P}_i : \mathbb{E}_{\mathbb{P}_i}[\xi_i] = \mu_i, \mathbb{E}_{\mathbb{P}_i}[(\xi_i - \mu_i)(\xi_i - \mu_i)^\top] = \Sigma_i \right\}, \quad (7.7)$$

where $\Sigma_i \succ 0$ for all $i \in [I]$.

Distributionally robust single chance constraints with moment based ambiguity sets as above have been considered in [33, 37].

Next we establish that, in general, it is strongly NP-hard to optimize over set Z_B . We will need the following result which shows that set Z_B can be recast as a bi-convex program. This confirms the statement in [73] that optimizing variables s_i in Bonferroni approximation “destroys the convexity.”

Lemma 12. *Under Assumption (A2), Z_B is equivalent to*

$$Z_B = \left\{ x : a_i(x)^\top \mu_i + \sqrt{\frac{1 - s_i}{s_i}} \sqrt{a_i(x)^\top \Sigma_i a_i(x)} \leq b_i(x), \forall i \in [I], \sum_{i \in [I]} s_i \leq \epsilon, s \geq 0 \right\}. \quad (7.8)$$

Proof. From [37] and [108], the chance constraint $\inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i\{\boldsymbol{\xi} : a_i(x)^\top \boldsymbol{\xi}_i \leq b_i(x)\} \geq 1 - s_i$ is equivalent to

$$a_i(x)^\top \mu_i + \sqrt{\frac{1-s_i}{s_i}} \sqrt{a_i(x)^\top \Sigma_i a_i(x)} \leq b_i(x)$$

for all $i \in [I]$. Then, the conclusion follows from the definition of Z_B . \square \square

Theorem 33. *It is strongly NP-hard to optimize over set Z_B .*

Proof. We prove by using a transformation from the feasibility problem of a binary program. First, we consider set $\bar{S} := \{x \in \{0, 1\}^n : Dx \geq d\}$, with given matrix $D \in \mathbb{Z}^{\tau \times n}$ and vector $d \in \mathbb{Z}^n$, and the following feasibility problem:

$$(\text{Binary Program}): \text{ Does there exist an } x \in \{0, 1\}^n \text{ such that } x \in \bar{S}? \quad (7.9)$$

Second, we consider an instance of Z_B with a projected ambiguity set in the form of (7.7) as

$$Z_B = \left\{ x : \begin{array}{l} \inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i \{ \boldsymbol{\xi}_i : \boldsymbol{\xi}_i x_i \leq x_i \sqrt{2n-1} \} \geq 1 - s_i, \forall i \in [n] \\ \inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i \{ \boldsymbol{\xi}_i : \boldsymbol{\xi}_i (1 - x_i) \leq (1 - x_i) \sqrt{2n-1} \} \geq 1 - s_i, \forall i \in [2n] \setminus [n] \\ \inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i \{ \boldsymbol{\xi}_i : 0 \leq D_{i-2n} x - d_{i-2n} \} \geq 1 - s_i, \forall i \in [2n + \tau] \setminus [2n] \\ \sum_{i \in [2n + \tau]} s_i \leq 0.5, \\ s \geq 0, \end{array} \right\}$$

where

$$\mathcal{D}_i = \{ \mathbb{P}_i : \mathbb{E}_{\mathbb{P}_i}[\boldsymbol{\xi}_i] = 0, \mathbb{E}_{\mathbb{P}_i}[\boldsymbol{\xi}_i^2] = 1 \}, \forall i \in [2n + \tau],$$

and D_j denotes the j th row of matrix D . It follows from Lemma 12 and Fourier-Motzkin elimination of variables $\{s_i\}_{i \in [2n + \tau] \setminus [2n]}$ that

$$Z_B = \left\{ x : \begin{array}{l} \sqrt{\frac{1-s_i}{s_i}} |x_i| \leq x_i \sqrt{2n-1}, \sqrt{\frac{1-s_{n+i}}{s_{n+i}}} |1-x_i| \leq (1-x_i) \sqrt{2n-1}, \forall i \in [n], \\ \sum_{i \in [2n]} s_i \leq 0.5, s \geq 0, Dx \geq d \end{array} \right\}.$$

It is clear that $x_i \in [0, 1]$ for all $x \in Z_B$. Then, by discussing whether $x_i > 0$ and $x_i < 1$ for each $i \in [n]$, we can further recast Z_B as

$$Z_B = \left\{ x : \begin{array}{l} s_i \geq \frac{1}{2n} \mathbb{I}(x_i > 0), s_{n+i} \geq \frac{1}{2n} \mathbb{I}(x_i < 1), \forall i \in [n], \\ \sum_{i \in [2n]} s_i \leq 0.5, s \geq 0, x \in [0, 1]^n, Dx \geq d \end{array} \right\}, \quad (7.10)$$

Third, for $x \in Z_B$, let $I_1 = \{i \in [n] : 1 > x_i > 0\}$, $I_2 = \{i \in [n] : x_i = 0\}$, and $I_3 = \{i \in [n] : x_i = 1\}$, where $|I_1| + |I_2| + |I_3| = n$. Then,

$$0.5 \geq \sum_{i \in [2n]} s_i \geq \sum_{i \in [n]} \left(\frac{1}{2n} \mathbb{I}(x_i > 0) + \frac{1}{2n} \mathbb{I}(x_i < 1) \right) = \frac{2|I_1| + |I_2| + |I_3|}{2n} = 0.5 + \frac{|I_1|}{2n},$$

where the first two inequalities are due to (7.10) and the third equality is due to the definitions of sets I_1 , I_2 , and I_3 . Hence, $|I_1| = 0$ and so $x \in \{0, 1\}^n$ for all $x \in Z_B$. It follows that $\bar{S} \supseteq Z_B$. On the other hand, for any $x \in \bar{S}$, by letting $s_i = \frac{1}{2n} \mathbb{I}(x_i > 0)$, $s_{n+i} = \frac{1}{2n} \mathbb{I}(x_i < 1)$, clearly, (x, s) satisfies (7.10). Thus, $\bar{S} = Z_B$, i.e., \bar{S} is feasible if and only if Z_B is feasible. Then, the conclusion follows from the strong NP-hardness of (Binary Program). \square \square

Although Z_B is in general computationally intractable, there exist important special cases where Z_B is convex and tractable. In the following theorems, we provide two sufficient conditions for the convexity of Z_B . The first condition requires a relatively small risk parameter ϵ and applies to the setting of uncertain constraint coefficients (i.e., $A^i \neq 0$ for some $i \in [I]$).

Theorem 34. *Suppose that Assumption (A2) holds and $B^i = 0$ for all $i \in [I]$ and $\epsilon \leq \frac{1}{1+(2\sqrt{\eta}+\sqrt{4\eta+3})^2}$, where $\eta = \max_{i \in [I]} \mu_i^\top \Sigma_i^{-1} \mu_i$. Then set Z_B is convex and is equivalent to*

$$Z_B = \left\{ x : a_i(x)^\top \mu_i \leq b^i, s_i \geq \frac{a_i(x)^\top \Sigma_i a_i(x)}{a_i(x)^\top \Sigma_i a_i(x) + (b^i - a_i(x)^\top \mu_i)^2}, \forall i \in [I], \sum_{i \in [I]} s_i \leq \epsilon, s \geq 0 \right\}. \quad (7.11)$$

Proof. First, $b_i(x) = b^i$ because $B^i = 0$ for all $i \in [I]$. The reformulation (7.11) follows from Lemma 12.

Hence, $a_i(x)^\top \Sigma_i a_i(x) / [a_i(x)^\top \Sigma_i a_i(x) + (b^i - a_i(x)^\top \mu_i)^2] \leq s_i \leq \epsilon \leq 1 / [1 + 1 + (2\sqrt{\eta} + \sqrt{4\eta + 3})^2]$. Since $(b^i - a_i(x)^\top \mu_i) \geq 0$, we have

$$\frac{b^i - a_i(x)^\top \mu_i}{\sqrt{a_i(x)^\top \Sigma_i a_i(x)}} \geq 2\sqrt{\eta} + \sqrt{4\eta + 3}. \quad (7.12a)$$

Hence, to show the convexity of Z_B , it suffices to show that the function $a_i(x)^\top \Sigma_i a_i(x) / [a_i(x)^\top \Sigma_i a_i(x) + (b^i - a_i(x)^\top \mu_i)^2]$ is convex when x satisfies (7.12a). To this end, we let $z_i := \Sigma_i^{1/2} a_i(x)$, $q_i := \Sigma_i^{-1/2} \mu_i$, and $k_i := (b^i - a_i(x)^\top \mu_i) / \sqrt{a_i(x)^\top \Sigma_i a_i(x)} = (b^i - q_i^\top z_i) / \sqrt{z_i^\top z_i}$. Then, $k_i \geq 2\sqrt{\eta} + \sqrt{4\eta + 3}$. Since $a_i(x)$ is affine in the variables x , it suffices to show that the function

$$f_i(z_i) = \frac{z_i^\top z_i}{z_i^\top z_i + (b^i - z_i^\top q_i)^2}$$

is convex in variables z_i when $k_i := (b^i - q_i^\top z_i) / \sqrt{z_i^\top z_i} \geq 2\sqrt{\eta} + \sqrt{4\eta + 3}$. To this end, we consider the Hessian of $f_i(z_i)$, denoted by $Hf_i(z_i)$, and show that $r^\top Hf_i(z_i)r \geq 0$ for an arbitrary $r \in \mathbb{R}^{m_i}$. Standard calculations yield

$$\begin{aligned} r^\top Hf_i(z_i)r &= 2 \left(z_i^\top z_i + (b^i - z_i^\top q_i)^2 \right)^{-3} \left\{ z_i^\top z_i \left[(b^i - z_i^\top q_i)^2 r^\top r - z_i^\top z_i (q_i^\top r)^2 \right. \right. \\ &\quad \left. \left. - 4(b^i - z_i^\top q_i)(q_i^\top r)(z_i^\top r) + 3(b^i - z_i^\top q_i)^2 (q_i^\top r)^2 \right] \right. \\ &\quad \left. + (b^i - z_i^\top q_i)^2 \left[r^\top r (b^i - z_i^\top q_i)^2 - 4(z_i^\top r)^2 + 4(b^i - z_i^\top q_i)(q_i^\top r)(z_i^\top r) \right] \right\} \\ &= 2 \left(z_i^\top z_i + (b^i - z_i^\top q_i)^2 \right)^{-3} \left[(k_i^4 + k_i^2)(z_i^\top z_i)^2 (r^\top r) - 4k_i^2 (z_i^\top z_i)(z_i^\top r)^2 \right. \\ &\quad \left. + (3k_i^2 - 1)(z_i^\top z_i)^2 (q_i^\top r)^2 + (4k_i^3 - 4k_i)(z_i^\top z_i)^{3/2} (q_i^\top r)(z_i^\top r) \right] \end{aligned} \quad (7.12b)$$

$$\begin{aligned} &\geq 2 \left(z_i^\top z_i + (b^i - z_i^\top q_i)^2 \right)^{-3} \left[(k_i^4 + k_i^2)(z_i^\top z_i)^2 (r^\top r) - 4k_i^2 (z_i^\top z_i)^2 (r^\top r) \right. \\ &\quad \left. - (4k_i^3 - 4k_i) \sqrt{q_i^\top q_i} (z_i^\top z_i)^2 (r^\top r) \right] \end{aligned} \quad (7.12c)$$

$$\geq 2 \left(z_i^\top z_i + (b^i - z_i^\top q_i)^2 \right)^{-3} (z_i^\top z_i)^2 (r^\top r) k_i^2 \left(k_i^2 - 4k_i \sqrt{q_i^\top q_i} - 3 \right) \quad (7.12d)$$

$$\geq 0 \quad (7.12e)$$

for all $r \in \mathbb{R}^{m_i}$. Above, equality (7.12b) is from the definition of k_i ; inequality (7.12c) follows from $3k_i^2 \geq 1$, $(4k_i^3 - 4k_i) \geq 0$ and the Cauchy-Schwarz inequalities $z_i^\top r \leq \sqrt{z_i^\top z_i} \sqrt{r^\top r}$ and

$q_i^\top r \leq \sqrt{q_i^\top q_i} \sqrt{r^\top r}$; inequality (7.12d) is due to the fact $k_i \geq 0$; and inequality (7.12e) is because $k_i \geq 2\sqrt{\eta} + \sqrt{4\eta + 3} \geq 2\sqrt{q_i^\top q_i} + \sqrt{4q_i^\top q_i + 3}$. \square \square

The second condition does not require a small risk parameter ϵ but is only applicable when the constraint coefficients are not affected by the uncertain parameters (right-hand side uncertainty), i.e. $A^i = 0$ for all $i \in [I]$.

Theorem 35. *Suppose that Assumption (A2) holds. Further assume that $A^i = 0$ for all $i \in [I]$ and $\epsilon \leq 0.75$. Then the set Z_B is convex and is equivalent to*

$$Z_B = \left\{ x : (a^i)^\top \mu_i + \sqrt{\frac{1-s_i}{s_i}} \sqrt{(a^i)^\top \Sigma_i a^i} \leq b_i(x), \forall i \in [I], \sum_{i \in [I]} s_i \leq \epsilon, s \geq 0 \right\}. \quad (7.13)$$

Proof. For all $i \in [I]$, $a_i(x) = a^i$ because $A^i = 0$. Thus, the reformulation (7.13) follows from Lemma 12. Hence, to show the convexity of Z_B , it suffices to show that function $\sqrt{(1-s_i)/s_i}$ is convex in s_i for $0 \leq s_i \leq \epsilon$. This follows from the fact that

$$\frac{d^2}{ds_i^2} \left(\sqrt{\frac{1-s_i}{s_i}} \right) = \frac{0.75 - s_i}{(1-s_i)^{3/2} s_i^{5/2}} \geq 0$$

because $0 \leq s_i \leq \epsilon \leq 0.75$. \square \square

The following example illustrate that Z_B is convex when condition of Theorem 34 holds and becomes non-convex when this condition does not hold.

Example 14. *Consider set Z_B with regard to a projected ambiguity set in the form of (7.7),*

$$Z_B = \left\{ x : \begin{array}{l} \inf_{\mathbb{P}_1 \in \mathcal{D}_1} \mathbb{P}_1 \{ \xi_1 : x_1 \xi_1 \leq 1 \} \geq 1 - s_1 \\ \inf_{\mathbb{P}_2 \in \mathcal{D}_2} \mathbb{P}_2 \{ \xi_2 : x_2 \xi_2 \leq 1 \} \geq 1 - s_2 \\ s_1 + s_2 \leq \epsilon \\ s_1, s_2 \geq 0 \end{array} \right\}$$

where

$$\mathcal{D}_1 = \{ \mathbb{P}_1 : \mathbb{E}_{\mathbb{P}_1}[\xi_1] = 0, \mathbb{E}_{\mathbb{P}_1}[\xi_1^2] = 1 \}, \mathcal{D}_2 = \{ \mathbb{P}_2 : \mathbb{E}_{\mathbb{P}_2}[\xi_2] = 0, \mathbb{E}_{\mathbb{P}_2}[\xi_2^2] = 1 \}$$

Projecting out variables s_1, s_2 yields

$$Z_B = \left\{ x \in \mathbb{R}^2 : \frac{x_1^2}{x_1^2 + 1} + \frac{x_2^2}{x_2^2 + 1} \leq \epsilon \right\}.$$

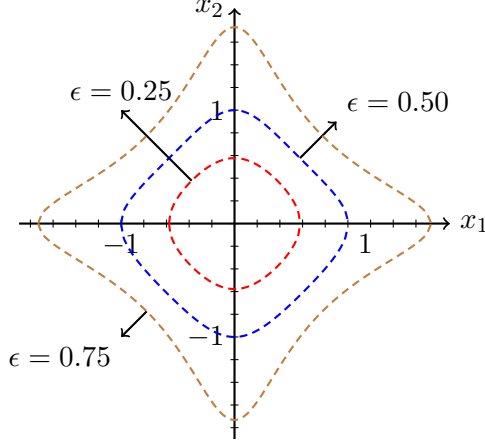


Figure 8: Illustration of Example 14

We depict Z_B in Figure 8 with $\epsilon = 0.25, 0.50$, and 0.75 , respectively, where the dashed line denotes the boundary of Z_B for each ϵ . Note that Z_B is convex when $\epsilon = 0.25$ and becomes non-convex when $\epsilon = 0.50, 0.75$. As $\eta = \max_{i \in [I]} \mu_i^\top \Sigma_i \mu_i = 0$, this figure confirms the sufficient condition of Theorem 34 that Z_B is convex when $\epsilon \leq \frac{1}{1+(2\sqrt{\eta}+\sqrt{4\eta+3})^2} = 0.25$.

□

Finally, we note that when either conditions of Theorem 34 or Theorem 35 hold, Z_B is not only convex but also computationally tractable. This observation follows from the classical result in [44] on the equivalence between tractable convex programming and the separation of a convex set from a point.

Theorem 36. *Under Assumption (A2), suppose that set S is closed and compact, and it is equipped with an oracle that can, for any $x \in \mathbb{R}^n$, either confirm $x \in S$ or provide a hyperplane that separates x from S in time polynomial in n . Additionally, suppose that either conditions of Theorem 34 or Theorem 35 holds. Then, for any $\alpha \in (0, 1)$, one can find an α -optimal solution to the optimized Bonferroni approximation of Z , i.e., formulation $\min_x \{c^\top x : x \in S \cap Z_B\}$, in time polynomial in $\log(1/\alpha)$ and the size of the formulation.*

Proof. We prove the conclusion when condition of Theorem 34 holds. The proof for the condition of Theorem 35 is similar and is omitted here for brevity.

Since S is convex by assumption and Z_B is convex by Theorem 34, the conclusion follows from Theorem 3.1 in [44] if we can show that there exists an oracle that can, for any $x \in \mathbb{R}^n$,

either confirm $x \in Z_B$ or provide a hyperplane that separates x from Z_B in time polynomial in n . To this end, from the proof of Theorem 34, we note that Z_B can be recast as

$$Z_B = \left\{ x : a_i(x)^\top \mu_i \leq b^i, \forall i \in [I], \sum_{i \in [I]} \frac{a_i(x)^\top \Sigma_i a_i(x)}{a_i(x)^\top \Sigma_i a_i(x) + (b^i - a_i(x)^\top \mu_i)^2} \leq \epsilon \right\}. \quad (7.15)$$

All constraints in (7.15) are linear except $\sum_{i \in [I]} g_i(x) \leq \epsilon$, where $g_i(x) := a_i(x)^\top \Sigma_i a_i(x) / [a_i(x)^\top \Sigma_i a_i(x) + (b^i - a_i(x)^\top \mu_i)^2]$. On one hand, whether or not $\sum_{i \in [I]} g_i(x) \leq \epsilon$ can be confirmed by a direct evaluation of $g_i(x)$, $i \in [I]$, in time polynomial in n . On the other hand, for an \hat{x} such that $\sum_{i \in [I]} g_i(\hat{x}) > \epsilon$, the following separating hyperplane can be obtained in time polynomial in n :

$$\epsilon \geq \sum_{i \in [I]} \left\{ g_i(\hat{x}) + \frac{2(b^i - q_i^\top \hat{z}_i)}{[\hat{z}_i^\top \hat{z}_i + (b^i - q_i^\top \hat{z}_i)^2]^2} \left[(b^i - q_i^\top \hat{z}_i) \hat{z}_i + (\hat{z}_i^\top \hat{z}_i) q_i \right]^\top \Sigma_i^{1/2} A^i (x - \hat{x}) \right\},$$

where $\hat{z}_i = \Sigma_i^{1/2} (A^i \hat{x} + a^i)$ and $q_i = \Sigma_i^{-1/2} \mu_i$. \square \square

7.4 Ambiguity Set Based on Marginal Distributions

In this section, we study the computational tractability of the optimized Bonferroni approximation when the ambiguity set is characterized by the (known) marginal distributions of the projected random vectors. More specifically, we make the following assumption on \mathcal{D}_i .

(A3) The projected ambiguity sets $\{\mathcal{D}_i\}_{i \in [I]}$ are characterized by the marginal distributions of ξ_i , i.e., $\mathcal{D}_i = \{\mathbb{P}_i\}$, where \mathbb{P}_i represents the probability distribution of ξ_i .

We first note that \mathcal{D}_i is a singleton for all $i \in [I]$ under Assumption (A3). By the definition of Bonferroni approximation (7.5), Z_B is equivalent to

$$Z_B = \left\{ x : \mathbb{P}_i \left\{ \xi_i : a_i(x)^\top \xi_i \leq b_i(x) \right\} \geq 1 - s_i, \forall i \in [I], \sum_{i \in [I]} s_i \leq \epsilon, s \geq 0 \right\}. \quad (7.16)$$

Next, we show that optimizing over Z_B in the form of (7.16) is computationally intractable. In particular, the corresponding optimization problem is strongly NP-hard even if $m_i = 1$, $A^i = 0$, and $a^i = 1$ for all $i \in [I]$, i.e., only right-hand side uncertainty.

Theorem 37. *Under Assumption (A3), suppose that $m_i = 1$, $A^i = 0$, and $a^i = 1$ for all $i \in [I]$. Then, it is strongly NP-hard to optimize over set Z_B .*

Proof. Similar to the proof of Theorem 33, we consider set $\bar{S} = \{x \in \{0, 1\}^n : Dx \geq d\}$, with given matrix $D \in \mathbb{Z}^{\tau \times n}$ and vector $d \in \mathbb{R}^n$, and show the reduction from (Binary Program) defined in (7.9). Second, we consider an instance of Z_B with a projected ambiguity set satisfying Assumption (A3) as

$$Z_B = \left\{ x : \begin{array}{l} \inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i \{ \xi_i : \xi_i \leq x_i \} \geq 1 - s_i, \forall i \in [n] \\ \inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i \{ \xi_i : \xi_i \leq (1 - x_i) \} \geq 1 - s_i, \forall i \in [2n] \setminus [n] \\ \inf_{\mathbb{P}_i \in \mathcal{D}_i} \mathbb{P}_i \{ \xi_i : 0 \leq D_{i-2n}x - d_{i-2n} \} \geq 1 - s_i, \forall i \in [2n + \tau] \setminus [2n] \\ \sum_{i \in [2n + \tau]} s_i \leq 0.5, \\ s \geq 0, \end{array} \right\}$$

where

$$\mathcal{D}_i = \{ \mathbb{P}_i : \xi \sim \mathcal{B}(1, 1/(2n)) \}, \forall i \in [2n + \tau],$$

and $\mathcal{B}(1, p)$ denotes Bernoulli distribution with probability of success equal to p . It follows from (7.16) and Fourier-Motzkin elimination of variables $\{s_i\}_{i \in [2n + \tau] \setminus [2n]}$ that

$$Z_B = \left\{ x : \begin{array}{l} s_i \geq \frac{1}{2n} \mathbb{I}(x_i < 1), s_{n+i} \geq \frac{1}{2n} \mathbb{I}(x_i > 0), \forall i \in [n], \\ \sum_{i \in [2n]} s_i \leq 0.5, s \geq 0, x \in [0, 1]^n, Dx \geq d \end{array} \right\}.$$

Following a similar proof as that of Theorem 33, we can show that $\bar{S} = Z_B$, i.e., \bar{S} is feasible if and only if Z_B is feasible. Then, the conclusion follows from the strong NP-hardness of (Binary Program) in (7.9). □ □

Next, we identify two important sufficient conditions where Z_B is convex. Similar to Theorem 34, the first condition holds for left-hand uncertain constraints with a small risk parameter ϵ .

Theorem 38. *Suppose that Assumption (A3) holds and $B^i = 0$ and $\xi_i \sim \mathcal{N}(\mu_i, \Sigma_i)$ for all $i \in [I]$ and $\epsilon \leq \frac{1}{2} - \frac{1}{2} \operatorname{erf}(\sqrt{\eta} + \sqrt{\eta + 0.75})$, where $\eta = \max_{i \in [I]} \mu_i^\top \Sigma_i^{-1} \mu_i$ and $\operatorname{erf}(\cdot), \operatorname{erf}^{-1}(\cdot)$ denote the error function and its inverse, respectively. Then the set Z_B is convex and is*

equivalent to

$$Z_B = \left\{ \begin{array}{l} a_i(x)^\top \mu_i \leq b^i, \forall i \in [I], \\ x : \frac{1}{1 + 2(\operatorname{erf}^{-1}(1 - 2s_i))^2} \geq \frac{a_i(x)^\top \Sigma_i a_i(x)}{a_i(x)^\top \Sigma_i a_i(x) + (b^i - a_i(x)^\top \mu_i)^2}, \forall i \in [I], \\ \sum_{i \in [I]} s_i \leq \epsilon, s \geq 0. \end{array} \right\} \quad (7.17)$$

Proof. First, $b_i(x) = b^i$ because $B^i = 0$ for all $i \in [I]$. Since $\xi_i \sim \mathcal{N}(\mu_i, \Sigma_i)$ for all $i \in [I]$, it follows from (7.16) that Z_B is equivalent to (7.17).

Let $f_i(x) := a_i(x)^\top \Sigma_i a_i(x) / [a_i(x)^\top \Sigma_i a_i(x) + (b^i - a_i(x)^\top \mu_i)^2]$. Since $\epsilon \leq \frac{1}{2} - \frac{1}{2} \operatorname{erf}(\sqrt{\eta} + \sqrt{\eta + 0.75})$ and $s_i \leq \epsilon$, thus we have $f_i(x) \leq 1/[1 + (2\sqrt{\eta} + \sqrt{4\eta + 3})^2]$. Hence, from the proof of Theorem 34, $f_i(x)$ is convex in $x \in Z_B$. Hence, it remains to show that $G(s_i) := 1/[1 + 2(\operatorname{erf}^{-1}(1 - 2s_i))^2]$ is concave in variable s_i when $s_i \in [0, \epsilon]$. This is indeed so because

$$\frac{d^2 G(s_i)}{ds_i^2} = -\frac{4\pi e^{2\operatorname{erf}^{-1}(1-2s_i)^2} [1 - 2\operatorname{erf}^{-1}(1 - 2s_i)^2]^2}{[1 + 2\operatorname{erf}^{-1}(1 - 2s_i)^2]^3} \leq 0$$

for all $0 \leq s_i \leq \epsilon$. □

□

Similar to Theorem 35, the second condition only holds for right-hand uncertain constraints with a relatively large risk parameter ϵ .

Theorem 39. Suppose that Assumption (A3) holds and $m_i = 1$, $A^i = 0$, $a^i = 1$ for all $i \in [I]$ and $\epsilon \leq \min_{i \in [I]} \{1 - F_i(r_i)\}$, where $F_i(\cdot)$ represents the cumulative distribution function of ξ_i and r_i represents its concave point, i.e., $F_i(r)$ is concave when $r \geq r_i$. Then the set Z_B is convex and is equivalent to

$$Z_B = \left\{ x : F_i(b_i(x)) \geq 1 - s_i, \forall i \in [I], \sum_{i \in [I]} s_i \leq \epsilon, s \geq 0 \right\}. \quad (7.18)$$

Proof. By assumption, ξ_i is a 1-dimensional random variable and so Z_B is equivalent to (7.18). Since $s_i \leq \epsilon$, $\epsilon \leq 1 - F_i(r_i)$ by assumption, and $b_i(x)$ is affine in x , it follows that the constraint $F_i(b_i(x)) \geq 1 - s_i$ is convex. Thus Z_B is convex. □ □

Table 11 displays some common probability distributions together with the concave points of their cumulative distribution function (cdf). Note that $1 - F(r^*)$, displayed in the last column of this table, represents an upper bound of ϵ in the condition of Theorem 39.

Table 11: Examples of Probability Distribution, cdf, and Concave Points

Distribution	cdf $F(r)$	Concave Point (r^*)	$1 - F(r^*)$
Normal $\mathcal{N}(\mu, \sigma^2)$	$\frac{1}{2} + \frac{1}{2}\text{erf}(\frac{r-\mu}{\sigma\sqrt{2}})$	μ	0.5
Exponential(λ)	$1 - \exp^{-\lambda r}, r \geq 0$	0	1
Uniform $[\ell, u]$	$\frac{r-\ell}{u-\ell}, \ell \leq r \leq u$	ℓ	1
Weibull(λ, k)	$1 - e^{-(r/\lambda)^k}, r \geq 0$	$\begin{cases} 0 \\ \lambda(k-1)^{1/k} \end{cases}$	$\begin{cases} 1, & \text{if } k \in (0, 1] \\ e^{1-k}, & \text{if } k > 1 \end{cases}$
Gamma(k, θ)	$1 - \frac{\Gamma(k, r/\theta)}{\Gamma(k)}, r \geq 0$	$\begin{cases} 0 \\ (k-1)\theta \end{cases}$	$\begin{cases} 1, & \text{if } k \in (0, 1] \\ \frac{\Gamma(k, k-1)}{\Gamma(k)}, & \text{if } k > 1 \end{cases}$
Log-Normal $\log \mathcal{N}(\mu, \sigma^2)$	$\frac{1}{2} + \frac{1}{2}\text{erf}(\frac{\log r - \mu}{\sigma\sqrt{2}}), r \geq 0$	$e^{\mu - \sigma^2}$	$\frac{1}{2} + \frac{1}{2}\text{erf}(\frac{\sigma}{\sqrt{2}})$
Logistic(μ, w)	$[1 + e^{-(r-\mu)/w}]^{-1}$	μ	0.5

Similar to Theorem 36, we note that when either the condition of Theorem 38 holds or that of Theorem 39 holds, the set Z_B is not only convex but also computationally tractable. We summarize this result in the following theorem and omit its proof.

Theorem 40. *Under Assumption (A3), suppose that set S is closed and compact, and it is equipped with an oracle that can, for any $x \in \mathbb{R}^n$, either confirm $x \in S$ or provide a hyperplane that separates x from S in time polynomial in n . Additionally, suppose that either condition of Theorem 38 or that of Theorem 39 holds. Then, for any $\alpha \in (0, 1)$, one can find an α -optimal solution to the problem $\min_x \{c^\top x : x \in S \cap Z_B\}$, in time polynomial in $\log(1/\alpha)$ and the size of the formulation.*

When modeling constraint uncertainty, besides the (parametric) probability distributions mentioned in Table 11, a nonparametric alternative employs the empirical distribution of ξ that can be directly established from the historical data. In the following theorem, we consider right-hand side uncertainty with discrete empirical distributions and show that the optimized Bonferroni approximation can be recast as a mixed-integer linear program (MILP).

Theorem 41. *Suppose that Assumption (A3) holds and $m_i = 1$, $A^i = 0$, and $a^i = 1$ for all $i \in [I]$. Additionally, suppose that $\mathbb{P}\{\xi_i = \xi_i^j\} = p_i^j$ for all $j \in [N_i]$ such that $\sum_{j \in [N_i]} p_i^j = 1$*

for all $i \in [I]$, and $\{\xi_i^j\}_{j \in [N_i]} \subset \mathbb{R}_+$ is sorted in the ascending order. Then,

$$Z_B = \left\{ x : \begin{array}{l} s_i \geq 0, z_i^j \in \{0, 1\}, \forall i \in [I], j \in [N_i], \\ \sum_{j \in [N_i]} \xi_i^j z_i^j \leq B^i x + b^i, \forall i \in [I], j \in [N_i], \\ \sum_{j \in [N_i]} \left(\sum_{t \in [j]} p_i^t \right) z_i^j \geq 1 - s_i, \forall i \in [I], j \in [N_i], \\ \sum_{j \in [N_i]} z_i^j = 1, \forall i \in [I], \\ \sum_{i \in [I]} s_i \leq \epsilon. \end{array} \right. \quad \begin{array}{l} (7.19a) \\ (7.19b) \\ (7.19c) \\ (7.19d) \\ (7.19e) \end{array}$$

Proof. By (7.16), $x \in Z_B$ if and only if there exists an $s_i \geq 0$ such that $\mathbb{P}_i\{\xi_i \leq B^i x + b^i\} \geq 1 - s_i$, $i \in [I]$, and $\sum_{i \in [I]} s_i \leq \epsilon$. Hence, it suffices to show that $\mathbb{P}_i\{\xi_i \leq B^i x + b^i\} \geq 1 - s_i$ is equivalent to constraints (7.19a)–(7.19d).

To this end, we note that nonnegative random variable ξ_i takes value ξ_i^j with probability p_i^j , and so $\mathbb{P}_i\{\xi_i \leq \xi_i^j\} = \sum_{t \in [j]} p_i^t$ for all $j \in [N_i]$. It follows that $\mathbb{P}_i\{\xi_i \leq B^i x + b^i\} \geq 1 - s_i$ holds if and only if $1 - s_i \leq \sum_{t \in [j]} p_i^t$ whenever $B^i x + b^i \geq \xi_i^j$. Then, we introduce additional binary variables $\{z_i^j\}_{j \in [N_i], i \in [I]}$ such that $z_i^j = 1$ when $B^i x + b^i \geq \xi_i^j$ and $z_i^j = 0$ otherwise. It follows that $\mathbb{P}_i\{\xi_i \leq B^i x + b^i\} \geq 1 - s_i$ is equivalent to constraints (7.19a)–(7.19d). $\square \quad \square$

Remark 6. The nonnegativity assumption of $\{\xi_i^j\}_{j \in [N_i]}$ for each $i \in [I]$ can be relaxed. If not, then for each $i \in [I]$ we can subtract M_i , where $M_i := \min_{j \in [N_i]} \xi_i^j$, from $\{\xi_i^j\}_{j \in [N_i]}$ and the right-hand side of uncertain constraint $B^i x + b^i$, i.e., $\xi_i^j := \xi_i^j - M_i$ for each $j \in [N_i]$ and $B^i x + b^i = B^i x + b^i - M_i$.

We close this section by demonstrating that Z_B may not be convex when the condition of Theorem 39 does not hold.

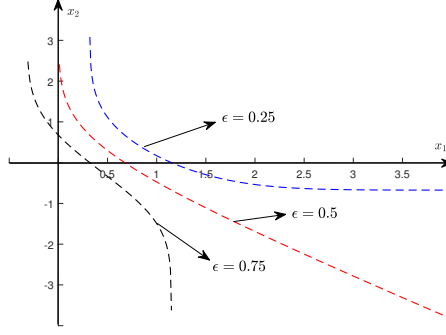


Figure 9: Illustration of Example 15

Example 15. Consider set Z_B with regard to a projected ambiguity set satisfying Assumption (A3),

$$Z_B = \left\{ x \in \mathbb{R}^2 : \begin{array}{l} \inf_{\mathbb{P}_1 \in \mathcal{D}_1} \mathbb{P}_1 \{ \xi_1 : \xi_1 \leq x_1 \} \geq 1 - s_1 \\ \inf_{\mathbb{P}_2 \in \mathcal{D}_2} \mathbb{P}_2 \{ \xi_2 : \xi_2 \leq x_1 \} \geq 1 - s_2 \\ \inf_{\mathbb{P}_3 \in \mathcal{D}_3} \mathbb{P}_3 \{ \xi_3 : \xi_3 \leq x_2 \} \geq 1 - s_3 \\ s_1 + s_2 + s_3 \leq \epsilon \\ s_1, s_2, s_3 \geq 0 \end{array} \right\}$$

where

$$\mathcal{D}_1 = \{ \mathbb{P}_1 : \xi_1 \sim \mathcal{N}(0, 1) \}, \mathcal{D}_2 = \{ \mathbb{P}_2 : \xi_2 \sim \mathcal{N}(0, 1) \}, \text{ and } \mathcal{D}_3 = \{ \mathbb{P}_3 : \xi_3 \sim \mathcal{N}(0, 1) \}$$

with standard normal distribution $\mathcal{N}(0, 1)$. Projecting out variables s_1, s_2, s_3 yields

$$Z_B = \left\{ x \in \mathbb{R}^2 : 2 \operatorname{erf} \left(\frac{x_1}{\sqrt{2}} \right) + \operatorname{erf} \left(\frac{x_2}{\sqrt{2}} \right) \geq 2 - 2\epsilon \right\}.$$

We depict Z_B in Fig. 9 with $\epsilon = 0.25, 0.50$, and 0.75 , respectively, where the dashed line denotes the boundary of Z_B for each ϵ . Note that this figure confirms condition of Theorem 39 that for normal random variables $\{\xi_i\}$, Z_B is convex if $\epsilon \leq 0.5$ but may not be convex otherwise. \square

7.5 Binary Decision Variables and Moment-based Ambiguity Sets

In this section, we focus on the projected ambiguity sets specified by first two moments as in Assumption (A2) and also assume that all decision variables x are binary, i.e.,

$S \subseteq \{0, 1\}^n$. Distributionally robust joint chance constrained optimization involving binary decision variables arise in a wide range of applications including the multi-knapsack problem (cf. [32, 118]) and the bin packing problem (cf. [101, 128]). In this case, Z_B is naturally non-convex due to the binary decision variables. Our goal, however, is to recast $S \cap Z_B$ as a mixed-integer second-order conic set (MSCS), which facilitates efficient computation with commercial solvers like GUROBI and CPLEX.

First, we show that $S \cap Z_B$ can be recast as an MSCS in the following result.

Theorem 42. *Under Assumption (A2), suppose that $S \subseteq \{0, 1\}^n$. Then, $S \cap Z_B = S \cap \widehat{Z}_B$, where*

$$\widehat{Z}_B = \left\{ x : \begin{array}{l} \mu_i^\top (A^i x + a^i) \leq B^i x + b^i, i \in [I], \\ \left\| \begin{bmatrix} 2\Sigma_i^{1/2}(A^i x + a^i) \\ s_i - t_i \end{bmatrix} \right\| \leq s_i + t_i, i \in [I], \\ t_i \leq \left(b^i - \mu_i^\top a^i \right)^2 + (a^i)^\top \Sigma_i a^i + 2 \left(b^i - \mu_i^\top a^i \right) \left(B^i - \mu_i^\top A^i \right) x \\ + 2(a^i)^\top \Sigma_i A^i x + \langle (B^i - \mu_i^\top A^i) (B^i - \mu_i^\top A^i)^\top + (A^i)^\top \Sigma_i A^i, w \rangle, i \in [I] \\ \sum_{i \in [I]} s_i \leq \epsilon, \\ w_{jk} \geq x_j + x_k - 1, 0 \leq w_{jk} \leq x_j, w_{jk} \leq x_k, \forall j, k \in [n], \\ s_i \geq 0, t_i \geq 0, \forall i \in [I]. \end{array} \right\} \quad (7.20)$$

Proof. By Lemma 12, we recast Z_B as

$$Z_B = \left\{ (x, y) : \begin{array}{l} a_i(x)^\top \mu_i \leq b_i(x), \\ a_i(x)^\top \Sigma_i a_i(x) \leq s_i \left[(b_i(x) - a_i(x)^\top \mu_i)^2 + a_i(x)^\top \Sigma_i a_i(x) \right], \forall i \in [I], \\ \sum_{i \in [I]} s_i \leq \epsilon, \\ s_i \geq 0, \forall i \in [I]. \end{array} \right\}$$

It is sufficient to linearize $(b_i(x) - a_i(x)^\top \mu_i)^2 + a_i(x)^\top \Sigma_i a_i(x)$ in the extended space for each $i \in [I]$. To achieve this, we introduce additional continuous variables $t_i := (b_i(x) - a_i(x)^\top \mu_i)^2 + a_i(x)^\top \Sigma_i a_i(x)$, $i \in [I]$, as well as additional binary variables $w := xx^\top$ and linearize them by using McCormick inequalities (see [71]), i.e.,

$$w_{jk} \geq x_j + x_k - 1, 0 \leq w_{jk} \leq x_j, w_{jk} \leq x_k, \forall j, k \in [n]$$

which lead to reformulation (7.20). \square \square

The reformulation of $S \cap Z_B$ in Theorem 42 incorporates n^2 auxiliary binary variables $\{w_{jk}\}_{j,k \in [n]}$. Next, under an additional assumption that $\epsilon \leq 0.25$, we show that it is possible to obtain a more compact reformulation by incorporating $n \times I$ auxiliary continuous variables when $I < n$.

Theorem 43. *Under Assumption (A2), suppose that $S \subseteq \{0, 1\}^n$ and $\epsilon \leq 0.25$. Then,*

$S \cap Z_B = S \cap \bar{Z}_B$, where

$$\bar{Z}_B = \left\{ x : \begin{array}{l} \mu_i^\top (A^i x + a^i) \leq B^i x + b^i, i \in [I], \\ \left\| \Sigma_i^{1/2} (A^i x + a^i) \right\| \leq (b^i - \mu_i^\top a^i) r_i + (B^i - \mu_i^\top A^i) q_{i\cdot}, \forall i \in [I], \\ \sum_{i \in [I]} s_i \leq \epsilon, \\ r_i \leq \sqrt{\frac{s_i}{1-s_i}}, \forall i \in [I], \\ q_{ij} \geq r_i - \sqrt{\frac{\epsilon}{1-\epsilon}} (1 - x_j), 0 \leq q_{ij} \leq \sqrt{\frac{\epsilon}{1-\epsilon}} x_j, q_{ij} \leq r_i, \forall i \in [I], j \in [n], \\ s_i \geq 0, r_i \geq 0, \forall i \in [I], \end{array} \right\} \quad (7.21a)$$

$$(7.21b)$$

$$(7.21c)$$

$$(7.21d)$$

$$(7.21e)$$

$$(7.21f)$$

where vector $q_{i\cdot} := [q_{i1}, \dots, q_{in}]^\top$ for all $i \in [I]$.

Proof. By Lemma 12, we recast Z_B as

$$Z_B = \left\{ (x, y) : \begin{array}{l} a_i(x)^\top \mu_i \leq b_i(x), \\ \sqrt{a_i(x)^\top \Sigma_i a_i(x)} \leq \sqrt{\frac{s_i}{1-s_i}} (b_i(x) - a_i(x)^\top \mu_i), \forall i \in [I], \\ \sum_{i \in [I]} s_i \leq \epsilon, \\ s_i \geq 0, \forall i \in [I]. \end{array} \right\} \quad (7.22a)$$

$$(7.22b)$$

$$(7.22c)$$

$$(7.22d)$$

We note that nonlinear constraints (7.22b) hold if and only if there exist $\{r_i\}_{i \in [I]}$ such that $0 \leq r_i \leq \sqrt{s_i/(1-s_i)}$ and $\sqrt{a_i(x)^\top \Sigma_i a_i(x)} \leq r_i (b_i(x) - a_i(x)^\top \mu_i)$ for all $i \in [I]$. Note that $s_i \in [0, \epsilon]$ and so $r_i \leq \sqrt{s_i/(1-s_i)} \leq \sqrt{\epsilon/(1-\epsilon)}$. Defining n -dimensional vectors $q_{i\cdot} := r_i x$, $i \in [I]$, we recast constraints (7.22b) as (7.21b), (7.21d)–(7.21f), where constraints (7.21e) are McCormick inequalities that linearize products $r_i x$. Note that constraints (7.21d)

characterize a convex feasible region because $0 \leq s_i \leq \epsilon \leq 0.25$ and so $\sqrt{s_i/(1-s_i)}$ is concave in s_i . \square \square

Remark 7. *When solving the optimized Bonferroni approximation as a mixed-integer convex program based on reformulation (7.21), we can incorporate the supporting hyperplanes of constraints (7.21d) as valid inequalities in a branch-and-cut algorithm. In particular, for given $\hat{s} \in [0, \epsilon]$, the supporting hyperplane at point $(\hat{s}, \sqrt{\hat{s}/(1-\hat{s})})$ is*

$$r_i \leq \left[\frac{1}{2} \hat{s}^{-1/2} (1 - \hat{s})^{-3/2} \right] s_i + \hat{s}^{1/2} (1 - \hat{s})^{-3/2} \left(\frac{1}{2} - \hat{s} \right). \quad (7.23a)$$

Remark 8. *We can construct inner and outer approximations of reformulation (7.21) by relaxing and restricting constraints (7.21d), respectively. More specifically, constraints (7.21d) imply $r_i \leq \sqrt{s_i/(1-\epsilon)}$ because $s_i \leq \epsilon$ for all $i \in [I]$. It follows that constraints (7.21d) imply the second-order conic constraints*

$$\left\| \begin{bmatrix} r_i \\ \frac{s_i - (1-\epsilon)}{2(1-\epsilon)} \end{bmatrix} \right\| \leq \frac{s_i + (1-\epsilon)}{2(1-\epsilon)}, \forall i \in [I]. \quad (7.23b)$$

In the branch-and-cut algorithm, we could start by relaxing constraints (7.21d) as (7.23b) and then iteratively incorporate valid inequalities in the form of (7.23a). In contrast to (7.23b), we can obtain a conservative approximation of constraints (7.21d) by noting that these constraints hold if $r_i \leq \sqrt{s_i}$. It follows that constraints (7.21d) are implied by the second-order conic constraints

$$\left\| \begin{bmatrix} r_i \\ \frac{s_i - 1}{2} \end{bmatrix} \right\| \leq \frac{s_i + 1}{2}, \forall i \in [I]. \quad (7.23c)$$

Hence, we obtain an inner approximation of Bonferroni approximation by replacing constraints (7.21d) with (7.23c).

7.6 Extension: Ambiguity Set with One Linking Constraint

In previous sections, we have shown that $Z = Z_B$ under the separability condition of Assumption (A1) and established several sufficient conditions under which the set Z_B is

convex. In this section, we demonstrate that these results may help establish new convexity results for the set Z even when the ambiguity set is not separable.

In this section, we consider an ambiguity set specified by means of random vectors $\{\boldsymbol{\xi}_i\}_{i \in [I]}$ and a bound on the overall deviation from mean. In particular, the ambiguity set is as follows.

(A4) The ambiguity set \mathcal{P} is given as

$$\mathcal{P} = \left\{ \mathbb{P} : \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] = \mu, \sum_{i \in [I]} \mathbb{E}_{\mathbb{P}}[\|\boldsymbol{\xi}_i - \mu_i\|] \leq \Delta \right\}. \quad (7.24)$$

Note that we can equivalently express \mathcal{P} as follows:

$$\mathcal{P} = \{ \mathbb{P} : \text{Proj}_i(\mathbb{P}) = \mathbb{P}_i \in \mathcal{D}_i(\delta_i), \forall i \in [I], \forall \delta \in \mathcal{K} \}, \quad (7.25a)$$

where $\mathcal{K} := \{ \delta : \delta \geq 0, \sum_{i \in [I]} \delta_i \leq \Delta \}$ and for each $i \in [I]$ and $\delta \in \mathcal{K}$. The marginal ambiguity sets $\{\mathcal{D}_i(\delta_i)\}_{i \in [I]}$ are defined as

$$\mathcal{D}_i(\delta_i) = \{ \mathbb{P} : \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}_i] = \mu_i, \mathbb{E}_{\mathbb{P}}[\|\boldsymbol{\xi}_i - \mu_i\|] \leq \delta_i \}, \quad (7.25b)$$

where $\Xi_i = \mathbb{R}^{m_i}$ for all $i \in [I]$.

The following theorem shows that under Assumption (A4), the set Z can be reformulated as a convex program.

Theorem 44. *Suppose that the ambiguity set \mathcal{P} is defined as (7.25a) and $\Xi = \prod_{i \in [I]} \Xi_i$, then the set Z is equivalent to*

$$Z = \left\{ x : \frac{\Delta}{2\epsilon} \|a_i(x)\|_* + a_i(x)^\top \mu_i \leq b_i(x), \forall i \in [I] \right\}, \quad (7.26)$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

Proof. We can reformulate Z as

$$Z = \{ x : x \in Z(\delta), \forall \delta \in \mathcal{K} \} \quad (7.27a)$$

where $\mathcal{K} := \{ \delta : \delta \geq 0, \sum_{i \in [I]} \delta_i \leq \Delta \}$ and

$$Z(\delta) := \left\{ x \in \mathbb{R}^n : \inf_{\mathbb{P} \in \mathcal{P}(\delta)} \mathbb{P} \left\{ \boldsymbol{\xi} : a_i(x)^\top \boldsymbol{\xi}_i \leq b_i(x), \forall i \in [I] \right\} \geq 1 - \epsilon \right\} \quad (7.27b)$$

with

$$\mathcal{P}(\delta) = \{\mathbb{P} : \text{Proj}_i(\mathbb{P}) = \mathbb{P}_i \in \mathcal{D}_i(\delta), \forall i \in [I]\}.$$

By Theorem 32, we know that $Z(\delta)$ is equivalent to its Bonferroni Approximation $Z_B(\delta)$ for any given $\delta \in \mathcal{K}$, i.e.,

$$Z(\delta) = Z_B(\delta) = \left\{ x : \inf_{\mathbb{P}_i \in \mathcal{D}_i(\delta_i)} \mathbb{P}_i \left\{ \xi_i : a_i(x)^\top \xi_i \leq b_i(x) \right\} \geq 1 - s_i, \forall i \in [I], \sum_{i \in [I]} s_i \leq \epsilon, s \geq 0 \right\}.$$

Let $\{\gamma_{1i}, \gamma_{2i}\}_{i \in [I]}$ be the dual variables corresponding to the moment constraints in (7.25b). Thus, by Theorem 4 in [118], set $Z_B(\delta)$ is equivalent to

$$Z_B(\delta) = \left\{ x : \begin{aligned} & \frac{1}{s_i} \gamma_{2i} \delta_i + \frac{(1 - s_i)}{s_i} \sup_{\xi_i} \left(\gamma_{1i}^\top (\xi_i - \mu_i) - \gamma_{2i} \|\xi_i - \mu_i\| \right) \\ & + \sup_{\xi_i} \left(\gamma_{1i}^\top (\xi_i - \mu_i) - \gamma_{2i} \|\xi_i - \mu_i\| - (b_i(x) - a_i(x)^\top \xi_i) \right) \leq 0, \forall i \in [I], \\ & \sum_{i \in [I]} s_i \leq \epsilon, \\ & \gamma_2 \geq 0, s \geq 0, \end{aligned} \right\}$$

where by convention, $0 \cdot \infty = 0$. By solving the inner supremums, $Z_B(\delta)$ is equivalent to

$$Z_B(\delta) = \left\{ x : \begin{aligned} & \frac{\gamma_{2i} \delta_i}{s_i} \leq b_i(x) - a_i(x)^\top \mu_i, \|\gamma_{1i}\|_* \leq \gamma_{2i}, \|\gamma_{1i} + a_i(x)\|_* \leq \gamma_{2i}, \forall i \in [I], \gamma_2 \geq 0, \\ & \sum_{i \in [I]} s_i \leq \epsilon, s \geq 0. \end{aligned} \right\} \quad (7.27c)$$

Now let

$$\tilde{Z}_B(\delta) = \left\{ x : \begin{aligned} & \frac{\gamma_{2i} \delta_i}{s_i} \leq b_i(x) - a_i(x)^\top \mu_i, \|a_i(x)\|_* \leq 2\gamma_{2i}, \forall i \in [I], \gamma_2 \geq 0, \\ & \sum_{i \in [I]} s_i \leq \epsilon, s \geq 0. \end{aligned} \right\} \quad (7.27d)$$

Note that $Z_B(\delta) \subseteq \tilde{Z}_B(\delta)$. This is because for each $i \in [I]$, by aggregating $\|\gamma_{1i}\|_* \leq \gamma_{2i}, \|\gamma_{1i} + a_i(x)\|_* \leq \gamma_{2i}$ and using triangle inequality, we have

$$\|a_i(x)\|_* \leq 2\gamma_{2i}.$$

On the other hand, by letting $\gamma_{1i} = -\frac{1}{2}a_i(x)$ in (7.27c), we obtain set $\tilde{Z}_B(\delta)$, thus $\tilde{Z}_B(\delta) \subseteq Z_B(\delta)$. Hence $\tilde{Z}_B(\delta) = Z_B(\delta)$.

By projecting out $\{\gamma_{2i}\}_{i \in [I]}$, (7.27d) yields

$$Z_B(\delta) = \left\{ x : \frac{\delta_i \|a_i(x)\|_*}{2s_i} \leq b_i(x) - a_i(x)^\top \mu_i, \forall i \in [I], \sum_{i \in [I]} s_i \leq \epsilon, s \geq 0 \right\}. \quad (7.27e)$$

Finally, by projecting out variables s , (7.27e) is further reduced to

$$Z_B(\delta) = \left\{ x : b_i(x) \geq a_i(x)^\top \mu_i, \forall i \in [I], \sum_{i \in [I]} \frac{\delta_i \|a_i(x)\|_*}{2(b_i(x) - a_i(x)^\top \mu)} \leq \epsilon \right\}.$$

Therefore,

$$Z = \left\{ x : b_i(x) \geq a_i(x)^\top \mu_i, \forall i \in [I], \sum_{i \in [I]} \frac{\delta_i \|a_i(x)\|_*}{2(b_i(x) - a_i(x)^\top \mu)} \leq \epsilon, \forall \delta \in \mathcal{K} \right\},$$

with $\mathcal{K} = \{\delta : \delta \geq 0, \sum_{i \in [I]} \delta_i \leq \Delta\}$, which is equivalent to

$$Z = \left\{ x : b_i(x) \geq a_i(x)^\top \mu_i, \forall i \in [I], \sum_{i \in [I]} \frac{\delta_i \|a_i(x)\|_*}{2(b_i(x) - a_i(x)^\top \mu)} \leq \epsilon, \forall \delta \in \mathbf{ext}(\mathcal{K}) \right\}, \quad (7.27f)$$

with $\mathbf{ext}(\mathcal{K}) := \{0\} \cup \{\Delta \mathbf{e}_i\}_{i \in [I]}$ denoting the set of extreme points of \mathcal{K} . Thus, (7.27f) leads to (7.26). \square \square

Remark 9. *The technique for proving Theorem 44 is quite general and may be applied to other settings. For example, if the ambiguity set \mathcal{P} is defined by known mean and sum of component-wise standard deviations, then we can reformulate Z as a second-order conic set.*

Next we consider the optimized Bonferroni approximation of Z .

Theorem 45. *Suppose that the ambiguity set \mathcal{P} is defined as (7.25a) and $\Xi = \prod_{i \in [I]} \Xi_i$, then the set Z_B is equivalent to*

$$Z_B = \left\{ x : \frac{\Delta}{2} \sum_{i \in [I]} \frac{\|a_i(x)\|_*}{b_i(x) - a_i(x)^\top \mu_i} \leq \epsilon, a_i(x)^\top \mu_i \leq b_i(x), \forall i \in [I] \right\}, \quad (7.28)$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

Proof. The optimized Bonferroni approximation of set Z is

$$Z_B = \left\{ x : \inf_{\mathbb{P}_j \in \mathcal{D}_j(\Delta)} \mathbb{P}_j \left\{ \boldsymbol{\xi}_j : a_j(x)^\top \boldsymbol{\xi}_j \leq b_j(x) \right\} \geq 1 - s_j, \forall j \in [I], \sum_{j \in [I]} s_j \leq \epsilon, s \geq 0 \right\}$$

i.e.,

$$Z_B = \left\{ x : \inf_{\mathbb{P}_j \in \mathcal{D}_j(\Delta)} \mathbb{P}_j \left\{ \boldsymbol{\xi}_j : a_j(x)^\top \boldsymbol{\xi}_j \leq b_j(x) \right\} \geq 1 - s_j, \forall j \in [I], \sum_{j \in [I]} s_j \leq \epsilon, s \geq 0 \right\}.$$

By letting $I = 1$ in Theorem 44, we know that $\inf_{\mathbb{P}_j \in \mathcal{D}_j(\Delta)} \mathbb{P}_j \left\{ \boldsymbol{\xi}_j : a_j(x)^\top \boldsymbol{\xi}_j \leq b_j(x) \right\} \geq 1 - s_j$ is equivalent to

$$\frac{\Delta}{2\epsilon} \|a_j(x)\|_* + a_j(x)^\top \mu_j \leq b_j(x)$$

for each $j \in [I]$. Thus, set Z_B is further equivalent to

$$Z_B = \left\{ x : \frac{\Delta}{2s_j} \|a_j(x)\|_* + a_j(x)^\top \mu_j \leq b_j(x), \forall j \in [I], \sum_{j \in [I]} s_j \leq \epsilon, s \geq 0 \right\},$$

which leads to (7.28) by projecting out s . □ □

Remark 10. *The constraints defining (7.28) are not convex in general. Thus even if Z is convex (Theorem 44), its optimized Bonferroni approximation Z_B may not be convex.*

Remark 11. *The constraints defining (7.28) are convex in case of only right-hand side uncertainties, i.e. $A^i = 0$ for all $i \in [I]$.*

We conclude by demonstrating the limitations of the optimized Bonferroni approximation by an example illustrating that, unless the established conditions hold, the distance between sets Z and Z_B can be arbitrarily large.

Example 16. *Consider Z with regard to a projected ambiguity set in the form of (7.25a)*

$$Z = \left\{ x \in \mathbb{R}^I : \inf_{\mathbb{P} \in \mathcal{P}} \mathcal{P} \left\{ \boldsymbol{\xi} : \boldsymbol{\xi}_i x_i \leq 1, \forall i \in [I] \right\} \geq 1 - \epsilon \right\}$$

where

$$\mathcal{P} = \{ \mathbb{P} : \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] = 0, \mathbb{E}_{\mathbb{P}}[\|\boldsymbol{\xi}\|] \leq \Delta \}.$$

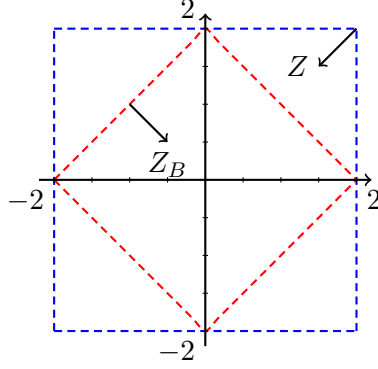


Figure 10: Illustration of Example 16 with $\frac{2\epsilon}{\Delta} = 2$ and $I = 2$

Thus, (7.26) and (7.28) yield

$$Z = \left\{ x \in \mathbb{R}^I : |x_i| \leq \frac{2\epsilon}{\Delta}, \forall i \in [I] \right\},$$

and

$$Z_B = \left\{ x \in \mathbb{R}^I : \sum_{i \in [I]} |x_i| \leq \frac{2\epsilon}{\Delta} \right\}.$$

These two sets are shown in Fig. 10 with $\frac{2\epsilon}{\Delta} = 2$ and $I = 2$, where the dashed lines denote the boundaries of Z, Z_B . Indeed, simple calculation shows that the Hausdorff distance (c.f. [88]) between sets Z_B and Z is $\frac{I-1}{\sqrt{I}} \frac{2\epsilon}{\Delta}$, which tends to be infinity when $\Delta \rightarrow 0$ and I, ϵ are fixed, or $I \rightarrow \infty$ and Δ, ϵ are fixed. \square

Bibliography

- [1] ABDI, A. and FUKASAWA, R., “On the mixing set with a knapsack constraint,” To appear in *Mathematical Programming*, 2016.
- [2] AHMED, S., “A scenario decomposition algorithm for 0-1 stochastic programs,” *Oper. Res. Lett.*, vol. 41, pp. 555–569, 2013.
- [3] AHMED, S., “Convex relaxations of chance constrained optimization problems,” *Optimization Letters*, vol. 8, no. 1, pp. 1–12, 2014.
- [4] AHMED, S., LUEDTKE, J., SONG, Y., and XIE, W., “Nonanticipative duality, relaxations, and formulations for chance-constrained stochastic programs,” *Mathematical Programming*, pp. 1–31, 2016.
- [5] AHMED, S. and PAPAGEORGIOU, D. J., “Probabilistic set covering with correlations,” *Operations Research*, vol. 61, no. 2, pp. 438–452, 2013.
- [6] ALON, N., ARORA, S., MANOKARAN, R., MOSHKOVITZ, D., and WEINSTEIN, O., “Inapproximability of densest κ -subgraph from average case hardness,” *Unpublished manuscript*, 2011.
- [7] ATAMTÜRK, A., NEMHAUSER, G. L., and SAVELSBERGH, M. W., “The mixed vertex packing problem,” *Mathematical Programming*, vol. 89, no. 1, pp. 35–53, 2000.
- [8] BAI, X., SUN, J., SUN, X., and ZHENG, X., “An alternating direction method for chance-constrained optimization problems with discrete distributions.” Available at http://www.optimization-online.org/DB_FILE/2012/04/3448.pdf, 2012.
- [9] BAKER, K., DALL’ANESE, E., and SUMMERS, T., “Distribution-agnostic stochastic optimal power flow for distribution grids,” in *North American Power Symposium (NAPS), 2016*, pp. 1–6, IEEE, 2016.
- [10] BALAS, E., “A modified lift-and-project procedure,” *Math. Program.*, vol. 79, pp. 19–31, 1997.
- [11] BEN-TAL, A., EL GHAOU, L., and NEMIROVSKI, A., *Robust optimization*. Princeton University Press, 2009.
- [12] BEN-TAL, A. and NEMIROVSKI, A., *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, vol. 2. SIAM, 2001.
- [13] BENNETT, K. P. and MANGASARIAN, O. L., “Bilinear separation of two sets inn-space,” *Computational Optimization and Applications*, vol. 2, no. 3, pp. 207–227, 1993.
- [14] BERALDI, P. and RUSZCZYŃSKI, A., “A branch and bound method for stochastic integer problems under probabilistic constraints,” *Optimization Methods and Software*, vol. 17, no. 3, pp. 359–382, 2002.
- [15] BERALDI, P. and RUSZCZYNSKI, A., “The probabilistic set-covering problem,” *Operations Research*, vol. 50, no. 6, pp. 956–967, 2002.

- [16] BERTSEKAS, D. P., “Constrained optimization and lagrange multiplier methods,” *Computer Science and Applied Mathematics, Boston: Academic Press, 1982*, vol. 1, 1982.
- [17] BERTSIMAS, D., LITVINOV, E., SUN, X. A., ZHAO, J., and ZHENG, T., “Adaptive robust optimization for the security constrained unit commitment problem,” *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 52–63, 2013.
- [18] BIAN, Q., XIN, H., WANG, Z., GAN, D., and WONG, K. P., “Distributionally robust solution to the reserve scheduling problem with partial information of wind power,” *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2822–2823, 2015.
- [19] BIENSTOCK, D., CHERTKOV, M., and HARNETT, S., “Chance-constrained optimal power flow: Risk-aware network control under uncertainty,” *SIAM Review*, vol. 56, no. 3, pp. 461–495, 2014.
- [20] BONFERRONI, C. E., *Teoria statistica delle classi e calcolo delle probabilità*. Libreria internazionale Seeber, 1936.
- [21] BONNANS, J. F. and SHAPIRO, A., *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2000.
- [22] CALAFIORE, G. and CAMPI, M. C., “Uncertain convex programs: randomized solutions and confidence levels,” *Mathematical Programming*, vol. 102, no. 1, pp. 25–46, 2005.
- [23] CALAFIORE, G. C. and CAMPI, M. C., “The scenario approach to robust control design,” *Automatic Control, IEEE Transactions on*, vol. 51, no. 5, pp. 742–753, 2006.
- [24] CALAFIORE, G. C. and EL GHAOU, L., “On distributionally robust chance-constrained linear programs,” *Journal of Optimization Theory and Applications*, vol. 130, no. 1, pp. 1–22, 2006.
- [25] CAMM, J. D., RATURI, A. S., and TSUBAKITANI, S., “Cutting big M down to size,” *Interfaces*, vol. 20, no. 5, pp. 61–66, 1990.
- [26] CARØE, C. C. and SCHULTZ, R., “Dual decomposition in stochastic integer programming,” *Oper. Res. Lett.*, pp. 37–45, 1999.
- [27] CERIA, S. and SOARES, J., “Convex programming for disjunctive convex optimization,” *Mathematical Programming*, vol. 86, no. 3, pp. 595–614, 1999.
- [28] CHARNES, A. and COOPER, W. W., “Deterministic equivalents for optimizing and satisficing under chance constraints,” *Operations research*, vol. 11, no. 1, pp. 18–39, 1963.
- [29] CHARNES, A., COOPER, W. W., and SYMONDS, G. H., “Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil,” *Management Science*, vol. 4, no. 3, pp. 235–263, 1958.
- [30] CHEN, W., SIM, M., SUN, J., and TEO, C.-P., “From cvar to uncertainty set: Implications in joint chance-constrained optimization,” *Operations research*, vol. 58, no. 2, pp. 470–485, 2010.

- [31] CHEN, Z. and SPOONER, E., “Grid power quality with variable speed wind turbines,” *IEEE Transactions on Energy Conversion*, vol. 16, no. 2, pp. 148–154, 2001.
- [32] CHENG, J., DELAGE, E., and LISSER, A., “Distributionally robust stochastic knapsack problem,” *SIAM Journal on Optimization*, vol. 24, no. 3, pp. 1485–1506, 2014.
- [33] DELAGE, E. and YE, Y., “Distributionally robust optimization under moment uncertainty with application to data-driven problems,” *Operations research*, vol. 58, no. 3, pp. 595–612, 2010.
- [34] DENTCHEVA, D. and RÖMISCH, W., “Duality gaps in nonconvex stochastic optimization,” *Math. Program.*, vol. 101, pp. 515–535, 2004.
- [35] DENTCHEVA, D., PRÉKOPA, A., and RUSZCZYNSKI, A., “Concavity and efficient points of discrete distributions in probabilistic programming,” *Mathematical Programming*, vol. 89, pp. 55–77, 2000.
- [36] DURRETT, R., *Probability: theory and examples*. Cambridge university press, 1996.
- [37] EL GHAOU, L., OKS, M., and OUSTRY, F., “Worst-case value-at-risk and robust portfolio optimization: A conic programming approach,” *Operations Research*, vol. 51, no. 4, pp. 543–556, 2003.
- [38] FISCHETTI, M., LODI, A., and TRAMONTANI, A., “On the separation of disjunctive cuts,” *Mathematical Programming*, vol. 128, pp. 205–230, 2011.
- [39] FRÉCHET, M., “Généralisation du théoreme des probabilités totales,” *Fundamenta mathematicae*, vol. 1, no. 25, pp. 379–387, 1935.
- [40] GOURTANI, A., XU, H., POZO, D., and NGUYEN, T.-D., “Robust unit commitment with n-1,” *Mathematical Methods of Operations Research*, vol. 83, no. 3, pp. 373–408, 2016.
- [41] GOYAL, V. and RAVI, R., “Approximation algorithms for robust covering problems with chance constraints.” Available at <http://repository.cmu.edu/cgi/viewcontent.cgi?article=1365&context=tepper>, 2008.
- [42] GOYAL, V. and RAVI, R., “A PTAS for the chance-constrained knapsack problem with random item sizes,” *Operations Research Letters*, vol. 38, no. 3, pp. 161–164, 2010.
- [43] GRANT, M., BOYD, S., and YE, Y., “CVX: Matlab software for disciplined convex programming,” 2008.
- [44] GRÖTSCHEL, M., LOVÁSZ, L., and SCHRIJVER, A., “The ellipsoid method and its consequences in combinatorial optimization,” *Combinatorica*, vol. 1, no. 2, pp. 169–197, 1981.
- [45] GÜNLÜK, O. and POCHET, Y., “Mixing mixed-integer inequalities,” *Mathematical Programming*, vol. 90, no. 3, pp. 429–457, 2001.
- [46] HANASUSANTO, G. A., ROITCH, V., KUHN, D., and WIESEMANN, W., “A distributionally robust perspective on uncertainty quantification and chance constrained programming,” *Mathematical Programming*, vol. 151, no. 1, pp. 35–62, 2015.

- [47] HANASUSANTO, G. A., ROITCH, V., KUHN, D., and WIESEMANN, W., “Ambiguous joint chance constraints under mean and dispersion information,” *Operations Research*, 2017.
- [48] HONG, L. J., YANG, Y., and ZHANG, L., “Sequential convex approximations to joint chance constrained programs: A monte carlo approach,” *Operations Research*, vol. 59, no. 3, pp. 617–630, 2011.
- [49] JABR, R. A., “Adjustable robust OPF with renewable energy sources,” *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4742–4751, 2013.
- [50] JABR, R. A., KARAKI, S., and KORBANE, J. A., “Robust multi-period opf with storage and renewables,” *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2790–2799, 2015.
- [51] JIANG, R. and GUAN, Y., “Data-driven chance constrained stochastic program,” *Mathematical Programming*, pp. 1–37, 2015.
- [52] KALL, P. and WALLACE, S. W., *Stochastic programming*. Chichester: John Wiley & Sons, 1994.
- [53] KATAOKA, S., “A stochastic programming model,” *Econometrica: Journal of the Econometric Society*, pp. 181–196, 1963.
- [54] KHACHIYAN, L., “The problem of calculating the volume of a polyhedron is enumerably hard,” *Russian Mathematical Surveys*, vol. 44, no. 3, pp. 199–200, 1989.
- [55] KIM, K.-L. H. S.-J. and BOYD, S., “Power allocation with outage probability specifications in wireless shadowed fading channels via geometric programming,” *RR, Information Systems Laboratory, Stanford University*, 2008.
- [56] KÜÇÜKYAVUZ, S., “On mixing sets arising in chance-constrained programming,” *Mathematical programming*, vol. 132, no. 1-2, pp. 31–56, 2012.
- [57] LAGOA, C. M., LI, X., and SZNAIER, M., “Probabilistically constrained linear programs and risk-adjusted controller design,” *SIAM Journal on Optimization*, vol. 15, no. 3, pp. 938–951, 2005.
- [58] LEJEUNE, M. A. and MARGOT, F., “Solving chance-constrained optimization problems with stochastic quadratic inequalities,” *Operations Research*, vol. 64, no. 4, pp. 939–957, 2016.
- [59] LI, B., JIANG, R., and MATHIEU, J. L., “Ambiguous risk constraints with moment and unimodality information.” Available at Optimization Online, 2016.
- [60] LI, B., JIANG, R., and MATHIEU, J. L., “Distributionally robust risk-constrained optimal power flow using moment and unimodality information,” in *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pp. 2425–2430, IEEE, 2016.
- [61] LIU, X., KÜÇÜKYAVUZ, S., and LUEDTKE, J., “Decomposition algorithms for two-stage chance-constrained programs,” *Mathematical Programming*, 2014.

- [62] LIU, X. and XU, W., “Economic load dispatch constrained by wind power availability: A here-and-now approach,” *IEEE Transactions on sustainable energy*, vol. 1, no. 1, pp. 2–9, 2010.
- [63] LU, X., MCELROY, M. B., and KIVILUOMA, J., “Global potential for wind-generated electricity,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 27, pp. 10933–10938, 2009.
- [64] LUBIN, M., BIENSTOCK, D., and VIELMA, J. P., “Two-sided linear chance constraints and extensions,” *arXiv preprint arXiv:1507.01995*, 2015.
- [65] LUBIN, M., DVORKIN, Y., and BACKHAUS, S., “A robust approach to chance constrained optimal power flow with renewable generation,” *arXiv preprint arXiv:1504.06011*, 2015.
- [66] LUEDTKE, J., “A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support,” *Mathematical Programming*, vol. 146, no. 1-2, pp. 219–244, 2014.
- [67] LUEDTKE, J. and AHMED, S., “A sample approximation approach for optimization with probabilistic constraints,” *SIAM Journal on Optimization*, vol. 19, no. 2, pp. 674–699, 2008.
- [68] LUEDTKE, J., AHMED, S., and NEMHAUSER, G. L., “An integer programming approach for linear programs with probabilistic constraints,” *Mathematical Programming*, vol. 122, no. 2, pp. 247–272, 2010.
- [69] LUO, C., BANAKAR, H., SHEN, B., and OOI, B.-T., “Strategies to smooth wind power fluctuations of wind turbine generator,” *IEEE Transactions on Energy Conversion*, vol. 22, no. 2, pp. 341–349, 2007.
- [70] MAGNANTI, T., SHAPIRO, J., and WAGNER, M., “Generalized linear programming solves the dual,” *Management science*, vol. 22, no. 11, pp. 1195–1203, 1976.
- [71] MCCORMICK, G. P., “Computability of global solutions to factorable nonconvex programs: Part I- convex underestimating problems,” *Mathematical programming*, vol. 10, no. 1, pp. 147–175, 1976.
- [72] MEYER, R. R., “On the existence of optimal solutions to integer and mixed-integer programming problems,” *Mathematical Programming*, vol. 7, no. 1, pp. 223–235, 1974.
- [73] NEMIROVSKI, A. and SHAPIRO, A., “Convex approximations of chance constrained programs,” *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 969–996, 2006.
- [74] NEMIROVSKI, A. and SHAPIRO, A., “Scenario approximations of chance constraints,” in *Probabilistic and randomized methods for design under uncertainty*, pp. 3–47, Springer, 2006.
- [75] OZTURK, U. A., MAZUMDAR, M., and NORMAN, B. A., “A solution to the stochastic unit commitment problem using chance constrained programming,” *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1589–1598, 2004.

- [76] PAGNONCELLI, B. K., AHMED, S., and SHAPIRO, A., “Sample average approximation method for chance constrained programming: theory and applications,” *Journal of optimization theory and applications*, vol. 142, no. 2, pp. 399–416, 2009.
- [77] PAGNONCELLI, B. K., AHMED, S., and SHAPIRO, A., “Computational study of a chance constrained portfolio selection problem,” *J. Optim. Theory Appl.*, vol. 142, no. 2, pp. 399–416, 2009.
- [78] PAPAVALIOU, A., OREN, S. S., and O’NEILL, R. P., “Reserve requirements for wind power integration: A scenario-based stochastic programming framework,” *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2197–2206, 2011.
- [79] POPESCU, I., “Robust mean-covariance solutions for stochastic optimization,” *Operations Research*, vol. 55, no. 1, pp. 98–112, 2007.
- [80] PRÉKOPA, A., *Stochastic Programming*. Springer, 1995.
- [81] PREKOPA, A., “On probabilistic constrained programming,” in *Proceedings of the Princeton symposium on mathematical programming*, pp. 113–138, Princeton University Press Princeton, NJ, 1970.
- [82] QIU, F., AHMED, S., DEY, S. S., and WOLSEY, L. A., “Covering linear programming with violations,” *INFORMS Journal on Computing*, vol. 26, no. 3, pp. 531–546, 2014.
- [83] ROALD, L., MISRA, S., CHERTKOV, M., and OTHERS, “Optimal power flow with weighted chance constraints and general policies for generation control,” in *2015 54th IEEE Conference on Decision and Control (CDC)*, pp. 6927–6933, IEEE, 2015.
- [84] ROALD, L., OLDEWURTEL, F., VAN PARYS, B., and ANDERSSON, G., “Security constrained optimal power flow with distributionally robust chance constraints,” *arXiv preprint arXiv:1508.06061*, 2015.
- [85] ROCKAFELLAR, R. T. and URYASEV, S., “Conditional value-at-risk for general loss distributions,” *Journal of Banking & Finance*, vol. 26, no. 7, pp. 1443–1471, 2002.
- [86] ROCKAFELLAR, R. T. and URYASEV, S., “Optimization of conditional value-at-risk,” *Journal of risk*, vol. 2, pp. 21–42, 2000.
- [87] ROCKAFELLAR, R. T. and WETS, R.-B., “Nonanticipativity and l_1 -martingales in stochastic optimization problems,” in *Stochastic Systems: Modeling, Identification and Optimization, II*, pp. 170–187, Springer, 1976.
- [88] ROCKAFELLAR, R. T. and WETS, R. J.-B., *Variational analysis*, vol. 317. Springer Science & Business Media, 1998.
- [89] RÜSCHENDORF, L., “Sharpness of Fréchet-bounds,” *Probability Theory and Related Fields*, vol. 57, no. 2, pp. 293–302, 1981.
- [90] RYAN, S. M., WETS, R. J.-B., WOODRUFF, D. L., SILVA-MONROY, C., and WATSON, J.-P., “Toward scalable, parallel progressive hedging for stochastic unit commitment,” in *2013 IEEE Power & Energy Society General Meeting*, pp. 1–5, IEEE, 2013.

- [91] SALINETTI, G. and WETS, R. J.-B., “On the convergence of sequences of convex sets in finite dimensions,” *SIAM review*, vol. 21, no. 1, pp. 18–33, 1979.
- [92] SCHNEIDER, R., *Convex bodies: the Brunn–Minkowski theory*, vol. 151. Cambridge University Press, 2013.
- [93] SHAPIRO, A. and AHMED, S., “On a class of minimax stochastic programs,” *SIAM Journal on Optimization*, vol. 14, no. 4, pp. 1237–1249, 2004.
- [94] SHAPIRO, A., DENTCHEVA, D., and RUSZCZYŃSKI, A., *Lectures on stochastic programming: modeling and theory*, vol. 9. SIAM, 2009.
- [95] SHAPIRO, A. and KLEYWEGT, A., “Minimax analysis of stochastic problems,” *Optimization Methods and Software*, vol. 17, no. 3, pp. 523–542, 2002.
- [96] SHERWOOD, L., “US solar market trends,” tech. rep., Interstate Renewable Energy Council, 2007.
- [97] SHIINA, T., “Numerical solution technique for joint chance-constrained programming problem: An application to electric power capacity expansion,” *Journal of the Operations Research Society of Japan*, vol. 42, no. 2, pp. 128–140, 1999.
- [98] SLATER, M., “Lagrange multipliers revisited,” tech. rep., Cowles Foundation for Research in Economics, Yale University, 1959.
- [99] SNYDER, L. V. and DASKIN, M. S., “Models for reliable supply chain network design,” in *Critical Infrastructure*, pp. 257–289, Springer, 2007.
- [100] SONG, Y. and LUEDTKE, J., “Branch-and-cut approaches for chance-constrained formulations of reliable network design problems,” *Mathematical Programming Computation*, vol. 5, no. 4, pp. 397–432, 2013.
- [101] SONG, Y., LUEDTKE, J. R., and KÜÇÜKYAVUZ, S., “Chance-constrained binary packing problems,” *INFORMS Journal on Computing*, 2014.
- [102] SUMMERS, T., WARRINGTON, J., MORARI, M., and LYGEROS, J., “Stochastic optimal power flow based on conditional value at risk and distributional robustness,” *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 116–125, 2015.
- [103] SUN, H., XU, H., and WANG, Y., “Asymptotic analysis of sample average approximation for stochastic optimization problems with joint chance constraints via conditional value at risk and difference of convex functions,” *Journal of Optimization Theory and Applications*, vol. 161, no. 1, pp. 257–284, 2014.
- [104] SWAMY, C., “Risk-averse stochastic optimization: probabilistically-constrained models and algorithms for black-box distributions,” in *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pp. 1627–1646, SIAM, 2011.
- [105] TAKYI, A. K. and LENCE, B. J., “Surface water quality management using a multiple-realization chance constraint method,” *Water Resources Research*, vol. 35, no. 5, pp. 1657–1670, 1999.

- [106] TALLURI, S., NARASIMHAN, R., and NAIR, A., "Vendor performance with supply risk: A chance-constrained dea approach," *International Journal of Production Economics*, vol. 100, no. 2, pp. 212–222, 2006.
- [107] TESTER, J. W., ANDERSON, B., BATCHELOR, A., BLACKWELL, D., DIPIPPO, R., DRAKE, E., GARNISH, J., LIVESAY, B., MOORE, M. C., NICHOLS, K., and OTHERS, "The future of geothermal energy: Impact of enhanced geothermal systems (EGS) on the United States in the 21st century," *Massachusetts Institute of Technology*, vol. 209, 2006.
- [108] WAGNER, M. R., "Stochastic 0–1 linear programming under limited distributional information," *Operations Research Letters*, vol. 36, no. 2, pp. 150–156, 2008.
- [109] WANG, J., SHAHIDEHPOUR, M., and LI, Z., "Security-constrained unit commitment with volatile wind power generation," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1319–1327, 2008.
- [110] WANG, Q., GUAN, Y., and WANG, J., "A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 206–215, 2012.
- [111] WANG, Z., BIAN, Q., XIN, H., and GAN, D., "A distributionally robust co-ordinated reserve scheduling model considering cvar-based wind power reserve requirements," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 2, pp. 625–636, 2016.
- [112] WATSON, J. P., WETS, R. J.-B., and WOODRUFF, D. L., "Scalable heuristics for a class of chance-constrained stochastic programs," *INFORMS J. Comput.*, vol. 22, pp. 543–554, 2010.
- [113] WEI, W., LIU, F., and MEI, S., "Distributionally robust co-optimization of energy and reserve dispatch," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 1, pp. 289–300, 2016.
- [114] WETS, R., "Stochastic programming: Solution techniques and approximation schemes," in *Mathematical Programming: State-of-the-Art 1982*, pp. 566–603, Springer, 1983.
- [115] WETS, R., "Stochastic programming," *Handbooks in Operations Research and Management Science*, vol. 1, pp. 573–629, 1989.
- [116] WIESEMAN, W., KUHN, D., and SIM, M., "Distributionally robust convex optimization," *Operations Research*, vol. 62, no. 6, pp. 1358–1376, 2014.
- [117] XIE, W. and AHMED, S., "Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation." Available at http://www.optimization-online.org/DB_FILE/2016/12/5759.pdf, 2016.
- [118] XIE, W. and AHMED, S., "On deterministic reformulation of distributionally robust joint chance-constrained problems." Available at http://www.optimization-online.org/DB_FILE/2016/09/5639.pdf, 2016.

- [119] XIE, W. and AHMED, S., “On quantile cuts and their closure for chance constrained optimization problems.” Available at http://www.optimization-online.org/DB_FILE/2016/03/5367.pdf, 2016.
- [120] XIE, W. and AHMED, S., “On the quantile cut closure of chance-constrained problems,” in *International Conference on Integer Programming and Combinatorial Optimization*, pp. 398–409, Springer International Publishing, 2016.
- [121] XIE, W., AHMED, S., and JIANG, R., “Optimized bonferroni approximations of distributionally robust joint chance constraints.” Available at Optimization Online, 2017.
- [122] XIONG, P., JIRUTITIJAROEN, P., and SINGH, C., “A distributionally robust optimization model for unit commitment considering uncertain wind power generation,” *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 39–49, 2017.
- [123] YANG, W. and XU, H., “Distributionally robust chance constraints for non-linear uncertainties,” *Mathematical Programming*, vol. 155, no. 1-2, pp. 231–265, 2016.
- [124] ZHANG, H. and LI, P., “Probabilistic analysis for optimal power flow under uncertainty,” *IET Generation, Transmission & Distribution*, vol. 4, no. 5, pp. 553–561, 2010.
- [125] ZHANG, H. and LI, P., “Chance constrained programming for optimal power flow under uncertainty,” *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2417–2424, 2011.
- [126] ZHANG, M., KÜÇÜKYAVUZ, S., and GOEL, S., “A branch-and-cut method for dynamic decision under joint chance constraints,” *Manage. Sci.*, 2014.
- [127] ZHANG, M., KÜÇÜKYAVUZ, S., and GOEL, S., “A branch-and-cut method for dynamic decision making under joint chance constraints,” *Management Science*, vol. 60, no. 5, pp. 1317–1333, 2014.
- [128] ZHANG, Y., JIANG, R., and SHEN, S., “Distributionally robust chance-constrained bin packing.” Available at Optimization Online, 2016.
- [129] ZHANG, Y., SHEN, S., and MATHIEU, J., “Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads,” *IEEE Transactions on Power Systems*, 2016.
- [130] ZHANG, Y., SHEN, S., and MATHIEU, J. L., “Data-driven optimization approaches for optimal power flow with uncertain reserves from load control,” in *2015 American Control Conference (ACC)*, pp. 3013–3018, IEEE, 2015.
- [131] ZHANG, Y. and GIANNAKIS, G. B., “Robust optimal power flow with wind integration using conditional value-at-risk,” in *Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on*, pp. 654–659, IEEE, 2013.
- [132] ZHAO, C. and GUAN, Y., “Data-driven stochastic unit commitment for integrating wind generation,” *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2587–2596, 2016.

- [133] ZHAO, M., HUANG, K., and ZENG, B., “Strong inequalities for chance–constrained program.” Available at http://www.optimization-online.org/DB_FILE/2014/11/4634.pdf, 2014.
- [134] ZIMMERMAN, R. D., MURILLO-SÁNCHEZ, C. E., and THOMAS, R. J., “MAT-POWER: Steady-state operations, planning, and analysis tools for power systems research and education,” *IEEE Transactions on power systems*, vol. 26, no. 1, pp. 12–19, 2011.
- [135] ZYMLER, S., KUHN, D., and RUSTEM, B., “Distributionally robust joint chance constraints with second-order moment information,” *Mathematical Programming*, vol. 137, no. 1-2, pp. 167–198, 2013.